S: 86055 Text 87240

Access DB# _____

SEARCH REQUEST FORM

CRFE

Scientific and Technical Information Center

Requester's Full Name: Lisa Y. Cook Examiner # : 77134 Date: 2/5/03
Art Unit: 1641 Phone Number 30 5-0808 Serial Number: 09/845,739
Mail Box and Bldg/Room Location: CM1-7E12 Results Format Preferred (circle): PAPER DISK E-MAIL
Office CM1 7B-17

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Biopolymer marker indicative of disease state having a molecular weight of 1793 daltons
Inventors (please provide full names): George Jackowski, Brad Thatcher, John Marshall, Tommy Yrees

Earliest Priority Filing Date: 4/30/01

For Sequence Searches Only *Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Sequence Search for: SKITHRIHWESASLL complement c3? 6

Seq. Id. No. 1

also Sequence utility in detecting congestive heart failure. Biopolymer marker/antibodies/assay.

also see attached claims + bib sheet.

Thanks,

LY Cook

****

STAFF USE ONLY

Searcher: D. Schreiber
Searcher Phone #: 308-4292
Searcher Location: CM1 6A03
Date Searcher Picked Up: _____
Date Completed: [redacted] 2/21
Searcher Prep & Review Time: 40 (Seq) | 56 (Text)
Clerical Prep Time: _____
Online Time: 6 | 67

Type of Search

NA Sequence (#) _____
AA Sequence (#) 2
Structure (#) _____
Bibliographic ✓
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN 364.18
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems _____
WWW/Internet _____
Other (specify) _____

PTO-1590 (8-01)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:11 ; Search time 82 Seconds
(without alignments)
24.375 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	100.0	1540	22	ABG25976	Novel human diagno
2	79	100.0	1592	18	AAW34623	Human C3 protein m
3	79	100.0	1635	18	AAW34624	Human C3 protein m
4	79	100.0	1657	18	AAW34629	Human C3 protein m
5	79	100.0	1661	18	AAW34625	Human C3 protein m
6	79	100.0	1663	17	AAR94028	Human C3 precursor
7	79	100.0	1663	17	AAR94029	Human modified C3
8	79	100.0	1663	17	AAR94030	Human modified C3
9	79	100.0	1663	18	AAW34619	Human C3 protein m
10	79	100.0	1663	18	AAW34620	Human C3 protein m
11	79	100.0	1663	18	AAW34621	Human C3 protein m
12	79	100.0	1663	18	AAW34627	Human C3 protein m
13	79	100.0	1663	18	AAW34628	Human C3 protein m
14	79	100.0	1663	18	AAW34630	Human C3 protein m
15	79	100.0	1663	18	AAW40988	Human C3 protein m
16	79	100.0	1663	18	AAW40989	Human C3 protein m
17	79	100.0	1663	18	AAW40990	Human C3 protein m
18	79	100.0	1663	18	AAW34606	Wild type human C3
19	79	100.0	1663	18	AAW34607	Human C3 protein m
20	79	100.0	1663	18	AAW34608	Human C3 protein m
21	79	100.0	1663	18	AAW34609	Human C3 protein m
22	79	100.0	1663	18	AAW34610	Human C3 protein m
23	79	100.0	1663	18	AAW34611	Human C3 protein m
24	79	100.0	1663	18	AAW34612	Human C3 protein m
25	79	100.0	1663	18	AAW34613	Human C3 protein m
26	79	100.0	1663	18	AAW34614	Human C3 protein m
27	79	100.0	1663	18	AAW34615	Human C3 protein m
28	79	100.0	1663	18	AAW34616	Human C3 protein m
29	79	100.0	1663	18	AAW34617	Human C3 protein m
30	79	100.0	1663	18	AAW34618	Human C3 protein m
31	79	100.0	1667	18	AAW34626	Human C3 protein m
32	79	100.0	1667	18	AAW34631	Human C3 protein m
33	43	54.4	280	22	ABB12430	Human bone marrow
34	42	53.2	411	22	ABB58617	Drosophila melanog
35	41	51.9	386	12	AAR11545	T. hyo 39 kD famil
36	41	51.9	541	23	ABP29330	Streptococcus poly
37	41	51.9	615	23	ABB54019	Lactococcus lactis
38	40	50.6	66	21	AAB23453	Soybean LLS1 prote
39	40	50.6	72	23	ABP10890	Human ORFX protein
40	40	50.6	74	22	AAU15863	Human novel secret
41	40	50.6	120	22	AAU45935	Propionibacterium
42	40	50.6	146	21	AAB33260	Pinus radiata tran
43	40	50.6	251	21	AAG15895	Arabidopsis thalia
44	40	50.6	420	22	ABB69833	Drosophila melanog
45	40	50.6	563	21	AAP01934	G. max COI1. Glyc

ALIGNMENTS

RESULT 1
ABG25976
ID ABG25976 standard; Protein; 1540 AA.
XX
AC ABG25976;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25967.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS90163.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

priority date 4/30/01

```
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 56335; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1540 AA;

  Query Match            100.0%;  Score 79;  DB 22;  Length 1540;
  Best Local Similarity  100.0%;  Pred. No. 0.00042;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 2
AAW34623
ID   AAW34623 standard; Protein; 1592 AA.
XX
AC   AAW34623;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-1.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1591
FT                   /note= "R1591T mutation"
FT   Misc-difference 1592
FT                   /note= "E1592N mutation"
FT   Misc-difference 1593
FT                   /note= "A1593Stop mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1592 AA;

  Query Match            100.0%;  Score 79;  DB 18;  Length 1592;
  Best Local Similarity  100.0%;  Pred. No. 0.00043;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 3
AAW34624
ID   AAW34624 standard; Protein; 1635 AA.
XX
AC   AAW34624;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-2.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1636
FT                   /note= "wild type E mutated to stop codon"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
```

```
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1635 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1635;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 4
AAW34629
ID   AAW34629 standard; Protein; 1657 AA.
XX
AC   AAW34629;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FR-2.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1638..1645
FT                   /note= "wild type residues QDEENQKQ mutated to SS"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1657 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1657;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 5
AAW34625
ID   AAW34625 standard; Protein; 1661 AA.
XX
AC   AAW34625;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-3.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
```

```
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1607..1614
FT                   /note= "wild type residues LSSDFWGE mutated to KEALQI"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;    97WO-GB00603.
XX
PR   19-NOV-1996;    96GB-0024028.
PR   07-MAR-1996;    96GB-0004865.
PR   07-JUN-1996;    96GB-0011896.
PR   08-JUL-1996;    96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1661 AA;

  Query Match            100.0%;  Score 79;  DB 18;  Length 1661;
  Best Local Similarity  100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASLL 15
           |||||||||||||||
Db    1305 SKITHRIHWESASLL 1319


RESULT 6
AAR94028
ID   AAR94028 standard; Protein; 1663 AA.
XX
AC   AAR94028;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human C3 precursor.
XX
KW   C3 protein; convertase; Factor I; Factor H; complement.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;    95WO-GB02121.
XX
PR   04-MAY-1995;    95GB-0009102.
PR   08-SEP-1994;    94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
DR   N-PSDB; AAT17738.
XX
PT   Mutant complement pathway protein forming stable C3 convertase -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Disclosure; Fig 1; 81pp; English.
XX
CC   Human C3 protein (AAR94028) was produced by expression of a cDNA
CC   sequence (AAT17738) isolated from a human liver cDNA library.
CC   C3 is a complement pathway protein that is suceptible to cleavage
CC   by Factor I and is also susceptible to the inhibitory action
CC   of Factor H.  Mutants of C3 (AAR94029 and AAR94030) have been
CC   produced by site-directed mutagenesis.  These mutants can be
CC   used to super-active the complement system, or to induce
CC   localised super-activation at a specific target to increase
CC   the target's sensitivity to complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match            100.0%;  Score 79;  DB 17;  Length 1663;
  Best Local Similarity  100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASLL 15
           |||||||||||||||
Db    1305 SKITHRIHWESASLL 1319


RESULT 7
AAR94029
ID   AAR94029 standard; Protein; 1663 AA.
XX
AC   AAR94029;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human modified C3 (R1303X).
XX
KW   C3 protein; convertase; Factor I; Factor H; complement;
KW   tumour; infection; therapy.
XX
OS   Synthetic.
XX
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
FT   Misc-difference 1303
FT                   /label= Glu, Gly, Gln
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;   95WO-GB02121.
XX
PR   04-MAY-1995;   95GB-0009102.
PR   08-SEP-1994;   94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
XX
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Claim 8; Fig 1; 81pp; English.
XX
CC   A modified human C3 protein (AAR94029) differs from the wild-type
CC   (AAR94028) by substitution of Arg-1303 by glutamic acid, glycine
CC   or glutamine.  It is obtained by site-directed mutagenesis of
CC   C3-encoding cDNA (AAT17738).  The modification results in improved
CC   resistance to cleavage by Factor I in comparison to wild-type C3.
CC   This allows the modified C3 to be used therapeutically to
CC   super-active the complement system or the increase a target's
CC   (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC   complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 17;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 8
AAR94030
ID   AAR94030 standard; Protein; 1663 AA.
XX
AC   AAR94030;
XX
DT   21-MAY-1996  (first entry)
XX
DE   Human modified C3 (D752G, E753S, D754G).
XX
KW   C3 protein; convertase; Factor I; Factor H; complement; tumour;
KW   infection; therapy.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Sig_peptide
FT   Protein         23..667
FT                   /note= "C3 beta chain"
FT   Peptide         668..671
FT                   /note= "amino acids 668-671 are removed when the
FT                          precursor is cleaved into the alpha and
FT                          beta chains"
FT   Protein         672..1663
FT                   /note= "C3 alpha chain"
XX
PN   WO9607738-A2.
XX
PD   14-MAR-1996.
XX
PF   08-SEP-1995;   95WO-GB02121.
XX
PR   04-MAY-1995;   95GB-0009102.
PR   08-SEP-1994;   94GB-0018147.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1996-171613/17.
XX
PT   Mutant complement pathway protein forming stable C3 convertase  -
PT   for generalised complement depletion or localised complement
PT   activation
XX
PS   Claim 11; Fig 1; 81pp; English.
XX
CC   A modified human C3 protein (AAR94030) differs from the wild-type
CC   (AAR94028) by substitution of Asp-Glu-Asp at positions 752-754 by
CC   Gly-Ser-Gly.  It is obtained by site-directed mutagenesis of
CC   C3-encoding cDNA (AAT17738).  The modification reduces the
CC   interaction of C3b/C3i with Factor H in comparison to wild-type
CC   C3.  This allows the modified C3 to be used therapeutically to
CC   super-active the complement system or the increase a target's
CC   (e.g. tumour, pathogen or virus-infected cell) sensitivity to
CC   complement-mediated destruction.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 17;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 9
AAW34619
ID   AAW34619 standard; Protein; 1663 AA.
XX
AC   AAW34619;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant DV-9.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1216
FT                   /note= "D1216G mutation"
FT   Misc-difference 1217
```

```
FT                   /note= "K1217E mutation"
FT   Misc-difference 1218
FT                   /note= "N1218D mutation"
FT   Misc-difference 1219
FT                   /note= "R1219H mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 14; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          ||||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 10
AAW34620
ID   AAW34620 standard; Protein; 1663 AA.
XX
AC   AAW34620;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant CV-4.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1260
FT                   /note= "R1260N mutation"
FT   Misc-difference 1264
FT                   /note= "G1264E mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 14; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          ||||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 11
AAW34621
ID   AAW34621 standard; Protein; 1663 AA.
```

```
XX
AC   AAW34621;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant RY-1.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1427
FT                   /note= "R1427Q mutation"
FT   Misc-difference 1431
FT                   /note= "K1431D mutation"
FT   Misc-difference 1433
FT                   /note= "E1433Q mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 14; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match           100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity  100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASLL 15
           |||||||||||||||
Db    1305 SKITHRIHWESASLL 1319


RESULT 12
AAW34627
ID   AAW34627 standard; Protein; 1663 AA.
XX
AC   AAW34627;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FT-5.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1661
FT                   /note= "C1661S mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
```

```
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 13
AAW34628
ID   AAW34628 standard; Protein; 1663 AA.
XX
AC   AAW34628;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FR-2.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1633
FT                   /note= "E1633R mutation"
FT   Misc-difference 1634
FT                   /note= "E1634D mutation"
FT   Misc-difference 1635
FT                   /note= "D1635T mutation"
FT   Misc-difference 1636
FT                   /note= "E1636T mutation"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 14
AAW34630
ID   AAW34630 standard; Protein; 1663 AA.
XX
AC   AAW34630;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant FR-3.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1638..1645
FT                   /note= "wild type residues QDEENQKQ mutated to RSTRQRAA"
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Example 17; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
```

```
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 15
AAW40988
ID   AAW40988 standard; Protein; 1663 AA.
XX
AC   AAW40988;
XX
DT   09-APR-1998  (first entry)
XX
DE   Human C3 protein mutant R1303X, R1320X.
XX
KW   Human; C3 protein; convertase; complement pathway protein; infection;
KW   down-regulation resistant C3 convertase; xenograft rejection; therapy;
KW   complement-mediated disease; autoimmune disease; leukaemia cell; tumour;
KW   complement-mediated response; MHC-mismatched lymphocyte; mutein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1303
FT                   /label= Glu, Tyr, Cys, Trp, Gln, Gly
FT   Misc-difference 1320
FT                   /label= Glu, Tyr, Cys, Trp, Gln, Gly
XX
PN   WO9732981-A1.
XX
PD   12-SEP-1997.
XX
PF   04-MAR-1997;   97WO-GB00603.
XX
PR   19-NOV-1996;   96GB-0024028.
PR   07-MAR-1996;   96GB-0004865.
PR   07-JUN-1996;   96GB-0011896.
PR   08-JUL-1996;   96GB-0014293.
XX
PA   (IMUT-) IMUTRAN LTD.
XX
PI   Farries TC,  Harrison RA;
XX
DR   WPI; 1997-457534/42.
XX
PT   Modified complement pathway protein that forms C3 convertase
PT   resistant to down-regulation - used to exhaust the complement
PT   pathway by super-activation, especially for preventing graft
PT   rejection, etc.
XX
PS   Claim 6; Page -; 123pp; English.
XX
CC   This sequence represents a mutated human C3 protein of the invention
CC   (see AAW34606 for wild type protein). This protein is a protein of the
CC   invention, and is a modified native complement pathway protein (A) that
CC   forms a down-regulation resistant C3 convertase. (A), their variants,
CC   fragments and conjugates are used to deplete levels of complement
CC   pathway proteins (by superactivation until one or more components are
CC   exhausted), specifically to prevent rejection of foreign material
CC   (particularly a xenograft) but also to prevent complement-mediated
CC   diseases resulting from (surgical) injury or antibody-antigen interaction
CC   in autoimmune disease, also to localise and/or amplify endogenous
CC   complement protein conversion and deposition at a specific site (e.g. a
CC   virus, infected cell or tumour, to increase sensitivity to
CC   complement-mediated responses; a particular application is eliminating
CC   any cancer cells left after surgical removal of a tumour). Also
CC   contemplated is ex vivo treatment, especially by passing blood through a
CC   matrix containing (A) (this may remove additional anaphylactic peptides
CC   and other inflammatory mediators) or killing of leukaemia cells or
CC   MHC-mismatched lymphocytes in extracted bone marrow. Since (A) is not
CC   inhibited by factor I, it can bind repeatedly to factor B (which is then
CC   inactivated), causing inactivation of the alternative pathway by
CC   consumption of factor B.
XX
SQ   Sequence   1663 AA;

  Query Match             100.0%;  Score 79;  DB 18;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


Search completed: February 21, 2003, 14:16:01
Job time : 84 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:12 ; Search time 30 Seconds
(without alignments)
14.711 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	100.0	1663	2	US-08-793-126-1	Sequence 1, Appli
2	79	100.0	1663	4	US-09-132-271-1	Sequence 1, Appli
3	79	100.0	1663	4	US-09-142-334-22	Sequence 22, Appl
4	37	46.8	267	4	US-09-449-218D-43	Sequence 43, Appl
5	37	46.8	272	2	US-08-887-997B-2	Sequence 2, Appli
6	37	46.8	700	4	US-09-413-814-68	Sequence 68, Appl
7	37	46.8	943	2	US-08-808-982-7	Sequence 7, Appli
8	37	46.8	943	4	US-09-306-902A-7	Sequence 7, Appli
9	36	45.6	93	1	US-08-839-710-3	Sequence 3, Appli
10	36	45.6	93	2	US-09-066-262-3	Sequence 3, Appli
11	36	45.6	150	4	US-09-605-785-707	Sequence 707, App
12	36	45.6	166	2	US-08-729-103-4	Sequence 4, Appli
13	36	45.6	625	4	US-08-959-004-10	Sequence 10, Appl
14	36	45.6	751	4	US-08-969-415-2	Sequence 2, Appli
15	36	45.6	844	1	US-07-646-537B-2	Sequence 2, Appli
16	36	45.6	937	3	US-09-005-180A-4	Sequence 4, Appli
17	35.5	44.9	24	4	US-09-082-279B-1199	Sequence 1199, Ap
18	35.5	44.9	24	4	US-09-315-304B-1199	Sequence 1199, Ap
19	35	44.3	117	6	5514582-15	Patent No. 5514582
20	35	44.3	165	2	US-08-401-530A-7	Sequence 7, Appli
21	35	44.3	165	2	US-08-729-103-3	Sequence 3, Appli
22	35	44.3	165	2	US-08-709-662-7	Sequence 7, Appli
23	35	44.3	233	4	US-09-214-631-7	Sequence 7, Appli
24	35	44.3	234	1	US-08-299-567-5	Sequence 5, Appli
25	35	44.3	238	1	US-08-240-124-2	Sequence 2, Appli
26	35	44.3	238	1	US-08-453-943-2	Sequence 2, Appli
27	35	44.3	238	2	US-09-057-121-2	Sequence 2, Appli
28	35	44.3	238	4	US-09-358-734-2	Sequence 2, Appli
29	35	44.3	332	4	US-09-134-001C-3977	Sequence 3977, Ap
30	35	44.3	556	3	US-08-501-572-1	Sequence 1, Appli
31	35	44.3	556	3	US-09-040-444-1	Sequence 1, Appli
32	35	44.3	566	4	US-09-491-522-7	Sequence 7, Appli
33	35	44.3	577	4	US-09-486-382B-2	Sequence 2, Appli
34	35	44.3	577	4	US-09-486-382B-13	Sequence 13, Appl
35	35	44.3	971	4	US-09-405-728-2	Sequence 2, Appli
36	35	44.3	1190	1	US-08-337-690A-2	Sequence 2, Appli
37	35	44.3	1190	4	US-09-048-887-2	Sequence 2, Appli
38	35	44.3	1205	4	US-09-491-522-11	Sequence 11, Appl
39	35	44.3	1211	4	US-09-491-522-5	Sequence 5, Appli
40	34	43.0	27	4	US-09-082-279B-1197	Sequence 1197, Ap
41	34	43.0	27	4	US-09-315-304B-1197	Sequence 1197, Ap
42	34	43.0	105	2	US-08-889-013C-6	Sequence 6, Appli
43	34	43.0	230	1	US-08-052-205-11	Sequence 11, Appl
44	34	43.0	230	1	US-08-595-974-11	Sequence 11, Appl
45	34	43.0	252	1	US-08-052-205-9	Sequence 9, Appli

ALIGNMENTS

```
RESULT 1
US-08-793-126-1
; Sequence 1, Application US/08793126
; Patent No. 5849297
; GENERAL INFORMATION:
;    APPLICANT: Harrison, Richard Alexander
;    APPLICANT: Farries, Charles Timothy
;    TITLE OF INVENTION: MODIFIED HUMAN C3 PROTEINS
;    NUMBER OF SEQUENCES: 2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: HALE AND DORR LLP
;      STREET: 60 State Street
;      CITY: Boston
;      STATE: MA
;      COUNTRY: United States of America
;      ZIP: 02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/793,126
;      FILING DATE: 07-FEB-1997
;      CLASSIFICATION: 424
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Baker, Hollie L.
;      REGISTRATION NUMBER: 31,321
;      REFERENCE/DOCKET NUMBER: 102286.377
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (617) 526-6000
;      TELEFAX: (617) 526-5000
;  INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1663 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-793-126-1

  Query Match 100.0%; Score 79; DB 2; Length 1663;
  Best Local Similarity 100.0%; Pred. No. 0.00014;
  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319
```

```
RESULT 2
US-09-132-271-1
; Sequence 1, Application US/09132271
; Patent No. 6221657
;  GENERAL INFORMATION:
;    APPLICANT:  Harrison, Richard Alexander
;    APPLICANT:  Farries, Charles Timothy
;    TITLE OF INVENTION:  MODIFIED HUMAN C3 PROTEINS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HALE AND DORR LLP
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  United States of America
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/132,271
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,126
;      FILING DATE:  07-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Hollie L.
;      REGISTRATION NUMBER:  31,321
;      REFERENCE/DOCKET NUMBER:  102286.377
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 526-6000
;      TELEFAX:  (617) 526-5000
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1663 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-132-271-1

  Query Match             100.0%;  Score 79;  DB 4;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASLL 15
           |||||||||||||||
Db    1305 SKITHRIHWESASLL 1319


RESULT 3
US-09-142-334-22
; Sequence 22, Application US/09142334
; Patent No. 6268485
; GENERAL INFORMATION:
;  APPLICANT: Farries, Timothy C.
;  APPLICANT: Harrison, Richard A.
;  TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase
;  FILE REFERENCE: 4-30443/A/IMU/PCT
;  CURRENT APPLICATION NUMBER: US/09/142,334
;  CURRENT FILING DATE: 1999-04-15
;  EARLIER APPLICATION NUMBER: PCT/GB97/00603
;  EARLIER FILING DATE: 1997-03-04
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1663
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-142-334-22

  Query Match             100.0%;  Score 79;  DB 4;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASLL 15
           |||||||||||||||
Db    1305 SKITHRIHWESASLL 1319


RESULT 4
US-09-449-218D-43
; Sequence 43, Application US/09449218D
; Patent No. 6395511
; GENERAL INFORMATION:
;  APPLICANT: Brunkow, Mary E.
;  APPLICANT:  Galas, David J.
;  APPLICANT:  Kovacevich, Brian
;  APPLICANT:  Mulligan, John T.
;  APPLICANT:  Paeper, Bryan W.
;  APPLICANT:  Van Ness, Jeffrey
;  APPLICANT:  Winkler, David G.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INCREASING
;  TITLE OF INVENTION:  BONE MINERALIZATION
;  FILE REFERENCE: 240083.508
;  CURRENT APPLICATION NUMBER: US/09/449,218D
;  CURRENT FILING DATE:  1999-11-24
;  NUMBER OF SEQ ID NOS: 45
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 43
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-449-218D-43

  Query Match             46.8%;  Score 37;  DB 4;  Length 267;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 THRIHWES 11
           :| :|||:
Db     154 SHEVHWET 161


RESULT 5
US-08-887-997B-2
; Sequence 2, Application US/08887997B
; Patent No. 5935852
;  GENERAL INFORMATION:
;    APPLICANT:  FOLLETTIE, MAXIMILLIAN
;    APPLICANT:  DeROBERTIS, EDWARD M.
;    TITLE OF INVENTION:  Mammalian Cerberus-Like Protein &
;    TITLE OF INVENTION:  Compositions
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,997B
;      FILING DATE:  03-JUL-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
```

```
;      REFERENCE/DOCKET NUMBER:  GI 5290
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  272 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-887-997B-2

  Query Match             46.8%;  Score 37;  DB 2;  Length 272;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 THRIHWES 11
          :| :|||:
Db    154 SHEVHWET 161


RESULT 6
US-09-413-814-68
; Sequence 68, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;. APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;. APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;. EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 700
;.  TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-68

  Query Match             46.8%;  Score 37;  DB 4;  Length 700;
  Best Local Similarity   60.0%;  Pred. No. 2.9e+02;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      6 RIHWESASLL 15
          |:||: | ||
Db    269 RLHWDWAQLL 278


RESULT 7
US-08-808-982-7
; Sequence 7, Application US/08808982
; Patent No. 5939271
;  GENERAL INFORMATION:
;    APPLICANT:  Tessier-Lavigne, Marc
;    APPLICANT:  Leonardo, E. David
;    APPLICANT:  Hink, Lindsay
;    APPLICANT:  Masu, Masayuki
;    APPLICANT:  Kazuko, Keino-Masu
;    TITLE OF INVENTION:  Netrin Receptors
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 BUSH STREET, SUITE 3200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/808,982
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UC96-217
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 343-4341
;      TELEFAX:  (415) 343-4342
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  943 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  peptide
US-08-808-982-7

  Query Match             46.8%;  Score 37;  DB 2;  Length 943;
  Best Local Similarity   57.1%;  Pred. No. 3.9e+02;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASL 14
          |: || :|  ||||
Db    514 SRDTHFLHLRSASL 527


RESULT 8
US-09-306-902A-7
; Sequence 7, Application US/09306902A
; Patent No. 6277585
;  GENERAL INFORMATION:
;       APPLICANT: Tessier-Lavigne, Marc
;                  Leonardo, E. David
;                  Hink, Lindsay
;                  Masu, Masayuki
;                  Kazuko, Keino-Masu
;       TITLE OF INVENTION: Netrin Receptors
;       NUMBER OF SEQUENCES: 9
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;            STREET: 268 BUSH STREET, SUITE 3200
;            CITY: SAN FRANCISCO
;            STATE: CALIFORNIA
;            COUNTRY: USA
;            ZIP: 94104
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/306,902A
;            FILING DATE: 07-May-1999
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: OSMAN, RICHARD A
;            REGISTRATION NUMBER: 36,627
;            REFERENCE/DOCKET NUMBER: UC96-217
```

```
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 343-4341
;             TELEFAX: (415) 343-4342
;    INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 943 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: not relevant
;             TOPOLOGY: not relevant
;        MOLECULE TYPE: peptide
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-306-902A-7

  Query Match             46.8%;  Score 37;  DB 4;  Length 943;
  Best Local Similarity   57.1%;  Pred. No. 3.9e+02;
  Matches    8;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASL 14
          |: || :|  ||||
Db    514 SRDTHFLHLRSASL 527


RESULT 9
US-08-839-710-3
; Sequence 3, Application US/08839710
; Patent No. 5776698
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Streeter, David G.
;    TITLE OF INVENTION:  NEW REGULATOR OF GENE
;    TITLE OF INVENTION:  TRANSCRIPTION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,710
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0220 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  93 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  202344
US-08-839-710-3

  Query Match             45.6%;  Score 36;  DB 1;  Length 93;
  Best Local Similarity   45.5%;  Pred. No. 53;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 THRIHWESASL 14
          :||: || | :
Db     20 SHRVTWEGAEV 30


RESULT 10
US-09-066-262-3
; Sequence 3, Application US/09066262
; Patent No. 5965706
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Streeter, David G.
;    TITLE OF INVENTION:  NEW REGULATOR OF GENE
;    TITLE OF INVENTION:  TRANSCRIPTION
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,262
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/839,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0220 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  93 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  202344
US-09-066-262-3

  Query Match             45.6%;  Score 36;  DB 2;  Length 93;
  Best Local Similarity   45.5%;  Pred. No. 53;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 THRIHWESASL 14
          :||: || | :
Db     20 SHRVTWEGAEV 30


RESULT 11
US-09-605-785-707
; Sequence 707, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Henderson, Robert A.
```

```
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C16
;  CURRENT APPLICATION NUMBER: US/09/605,785
;  CURRENT FILING DATE:  2000-06-27
;  NUMBER OF SEQ ID NOS: 835
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 707
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-605-785-707

  Query Match             45.6%;  Score 36;  DB 4;  Length 150;
  Best Local Similarity   50.0%;  Pred. No. 86;
  Matches    5;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 ITHRIHWESA 12
          : || || :|
Db    133 LAHRRHWRNA 142


RESULT 12
US-08-729-103-4
; Sequence 4, Application US/08729103
; Patent No. 5837841
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN REG PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,103
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0138 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  166 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  474306
US-08-729-103-4

  Query Match             45.6%;  Score 36;  DB 2;  Length 166;
  Best Local Similarity   50.0%;  Pred. No. 96;
  Matches    7;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 KITHRIHWESASLL 15
            |   | || | ||:
Db    106 KKNRRWHWSSGSLV 119


RESULT 13
US-08-959-004-10
; Sequence 10, Application US/08959004
; Patent No. 6197543
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Kaser, Matthew
;    TITLE OF INVENTION:  HUMAN VESICLE MEMBRANE PROTEIN-LIKE
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/959,004
;      FILING DATE:  Herewith
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0414 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  625 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1665777
US-08-959-004-10

  Query Match             45.6%;  Score 36;  DB 4;  Length 625;
  Best Local Similarity   36.4%;  Pred. No. 3.7e+02;
  Matches    4;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 THRIHWESASL 14
          |: :||| : :
```

```
Db    234 TYSVHWEESDI 244


RESULT 14
US-08-969-415-2
; Sequence 2, Application US/08969415
; Patent No. 6410303
;  GENERAL INFORMATION:
;    APPLICANT:  TAKANO, Hiroyuki
;    APPLICANT:  HINO, Akihiro
;    APPLICANT:  IYO, Chie
;    APPLICANT:  SUZUKI, Yasuo
;    APPLICANT:  NAKAJIMA, Ryoichi
;    TITLE OF INVENTION:  FROZEN DOUGH-RESISTANT, PRACTICAL
;    TITLE OF INVENTION:  BAKER'S YEAST
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  419 7th Street N.W., Ste. 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/969,415
;      FILING DATE:  21-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 08-297886
;      FILING DATE:  23-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEIMARK, Sheridan
;      REGISTRATION NUMBER:  20,520
;      REFERENCE/DOCKET NUMBER:  TAKANO-9
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 628-5197
;      TELEFAX:  (202) 737-3528
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  751 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-969-415-2

  Query Match             45.6%;  Score 36;  DB 4;  Length 751;
  Best Local Similarity   53.8%;  Pred. No. 4.5e+02;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESAS 13
          :|| ||  :|||:
Db    560 TKIKHRTSYESAT 572


RESULT 15
US-07-646-537B-2
; Sequence 2, Application US/07646537B
; Patent No. 5348864
;  GENERAL INFORMATION:
;    APPLICANT:  Barbacid, Mariano
;    TITLE OF INVENTION:  Vav Proto-Oncogene Protein
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bristol-Myers Squibb Company
;      STREET:  P.O. Box 4000
;      CITY:  Princeton
;      STATE:  New Jersey
;      COUNTRY:  U.S.A.
;      ZIP:  08543-4000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/646,537B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaul, Timothy J.
;      REGISTRATION NUMBER:  33,111
;      REFERENCE/DOCKET NUMBER:  DC10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (609) 921-5901
;      TELEFAX:  (609) 921-4526
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  844 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-646-537B-2

  Query Match             45.6%;  Score 36;  DB 1;  Length 844;
  Best Local Similarity   45.5%;  Pred. No. 5.1e+02;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      4 THRIHWESASL 14
          :||: || | :
Db     20 SHRVTWEGAQV 30


Search completed: February 21, 2003, 14:18:10
Job time : 32 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:40 ; Search time 12 Seconds
(without alignments)
38.837 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 156504

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	100.0	16	9	US-09-846-345-1	Sequence 1, Appli
2	79	100.0	17	9	US-09-846-346-1	Sequence 1, Appli
3	79	100.0	1663	10	US-09-875-519A-22	Sequence 22, Appli
4	75	94.9	14	9	US-09-845-730-1	Sequence 1, Appli
5	66	83.5	12	9	US-09-846-349-1	Sequence 1, Appli
6	61	77.2	11	9	US-09-845-715-1	Sequence 1, Appli
7	53	67.1	10	9	US-09-845-731-1	Sequence 1, Appli
8	42	53.2	91	10	US-09-867-550-910	Sequence 910, App
9	40	50.6	74	10	US-09-764-864-816	Sequence 816, App
10	39	49.4	64	10	US-09-867-550-908	Sequence 908, App
11	38	48.1	2012	9	US-09-808-602-68	Sequence 68, Appl
12	37	46.8	40	10	US-09-864-761-35988	Sequence 35988, A
13	37	46.8	69	10	US-09-864-761-44965	Sequence 44965, A
14	37	46.8	103	10	US-09-864-761-34487	Sequence 34487, A
15	37	46.8	134	10	US-09-864-761-46114	Sequence 46114, A
16	37	46.8	267	9	US-10-044-716-12	Sequence 12, Appl
17	37	46.8	267	9	US-09-089-818B-8	Sequence 8, Appli
18	37	46.8	272	9	US-09-887-552A-2	Sequence 2, Appli
19	37	46.8	272	9	US-09-089-818B-2	Sequence 2, Appli
20	37	46.8	385	12	US-10-139-262-2	Sequence 2, Appli
21	37	46.8	433	12	US-10-139-262-6	Sequence 6, Appli
22	37	46.8	440	12	US-10-139-262-4	Sequence 4, Appli
23	37	46.8	451	10	US-09-938-330-2	Sequence 2, Appli
24	37	46.8	486	10	US-09-938-330-6	Sequence 6, Appli
25	37	46.8	609	10	US-09-815-242-11758	Sequence 11758, A
26	37	46.8	1156	12	US-10-014-070-5	Sequence 5, Appli
27	37	46.8	1213	10	US-09-938-330-14	Sequence 14, Appl
28	37	46.8	1216	10	US-09-938-330-12	Sequence 12, Appl
29	37	46.8	1219	10	US-09-938-330-10	Sequence 10, Appl
30	37	46.8	1222	10	US-09-938-330-8	Sequence 8, Appli
31	37	46.8	1223	12	US-10-014-070-2	Sequence 2, Appli
32	37	46.8	1232	10	US-09-938-330-18	Sequence 18, Appl
33	37	46.8	1235	10	US-09-938-330-16	Sequence 16, Appl
34	37	46.8	1249	10	US-09-938-330-22	Sequence 22, Appl
35	37	46.8	1252	10	US-09-938-330-20	Sequence 20, Appl
36	36	45.6	19	10	US-09-864-761-37939	Sequence 37939, A
37	36	45.6	26	10	US-09-864-761-34810	Sequence 34810, A
38	36	45.6	150	9	US-10-012-896-707	Sequence 707, App
39	36	45.6	150	9	US-09-895-793-707	Sequence 707, App
40	36	45.6	150	9	US-09-895-814-707	Sequence 707, App
41	36	45.6	150	10	US-09-759-143-707	Sequence 707, App
42	36	45.6	150	10	US-09-780-669-707	Sequence 707, App
43	36	45.6	150	10	US-09-822-827-707	Sequence 707, App
44	36	45.6	174	10	US-09-925-301-1182	Sequence 1182, Ap
45	36	45.6	174	10	US-09-925-297-753	Sequence 753, App

ALIGNMENTS

check! = patent # 09/846,345 6617308

RESULT 1
US-09-846-345-1
; Sequence 1, Application US/09846345
; Patent No. US20020161182A1
; GENERAL INFORMATION:
; APPLICANT: Jackowski, George
; TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECU
; TITLE OF INVENTION: OF 1865 DALTONS
; FILE REFERENCE: 2132.045
; CURRENT APPLICATION NUMBER: US/09/846,345
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-846-345-1

Query Match 100.0%; Score 79; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SKITHRIHWESASLL 15
||||||||||||||||
Db 2 SKITHRIHWESASLL 16

check! Non-final Gail ask for claims

RESULT 2
US-09-846-346-1
; Sequence 1, Application US/09846346
; Patent No. US20020160532A1
; GENERAL INFORMATION:
; APPLICANT: Jackowski, George
; TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECU
; TITLE OF INVENTION: OF 1998 DALTONS
; FILE REFERENCE: 2132.013
; CURRENT APPLICATION NUMBER: US/09/846,346
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1

```
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-346-1

  Query Match             100.0%;  Score 79;  DB 9;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 9e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db      2 SKITHRIHWESASLL 16


RESULT 3
US-09-875-519A-22
; Sequence 22, Application US/09875519A
; Patent No. US20020068059A1
; GENERAL INFORMATION:
;  APPLICANT: Farries, Timothy C.
;  APPLICANT:  Harrison, Richard A.
;  TITLE OF INVENTION: Down-Regulation Resistant C3 Convertase
;  FILE REFERENCE: 4-30443/A/IMU/PCT
;  CURRENT APPLICATION NUMBER: US/09/875,519A
;  CURRENT FILING DATE:  2001-06-06
;  PRIOR APPLICATION NUMBER: PCT/GB97/00603
;  PRIOR FILING DATE: 1997-03-04
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1663
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-875-519A-22

  Query Match             100.0%;  Score 79;  DB 10;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 9e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 4
US-09-845-730-1
; Sequence 1, Application US/09845730
; Patent No. US20020169278A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECULAR
;  TITLE OF INVENTION:  OF 1690 DALTONS
;  FILE REFERENCE: 2132.042
;  CURRENT APPLICATION NUMBER: US/09/845,730
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 14
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-730-1

  Query Match             94.9%;  Score 75;  DB 9;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 KITHRIHWESASLL 15
          ||||||||||||||
Db      1 KITHRIHWESASLL 14


RESULT 5
US-09-846-349-1
; Sequence 1, Application US/09846349
; Patent No. US20020161186A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECUL
;  TITLE OF INVENTION:  OF 1449 DALTONS
;  FILE REFERENCE: 2132.034
;  CURRENT APPLICATION NUMBER: US/09/846,349
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 12
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-846-349-1

  Query Match             83.5%;  Score 66;  DB 9;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-05;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 THRIHWESASLL 15
          ||||||||||||
Db      1 THRIHWESASLL 12


RESULT 6
US-09-845-715-1
; Sequence 1, Application US/09845715
; Patent No. US20020161184A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: BIOPOLYMER MARKER INDICATIVE OF DISEASE STATE HAVING A MOLECUL
;  TITLE OF INVENTION:  OF 1348 DALTONS
;  FILE REFERENCE: 2132.030
;  CURRENT APPLICATION NUMBER: US/09/845,715
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 11
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-715-1

  Query Match             77.2%;  Score 61;  DB 9;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 0.00044;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
          |||||||||||
Db      1 HRIHWESASLL 11


RESULT 7
US-09-845-731-1
; Sequence 1, Application US/09845731
; Publication No. US20030004307A1
; GENERAL INFORMATION:
;  APPLICANT: Jackowski, George
;  TITLE OF INVENTION: Biopolymer Marker Indicative Of Disease State Having A Molecul
;  TITLE OF INVENTION:  Of 1211 Daltons
;  FILE REFERENCE: 2132.029
;  CURRENT APPLICATION NUMBER: US/09/845,731
;  CURRENT FILING DATE:  2001-04-30
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-845-731-1
```

```
  Query Match             67.1%;  Score 53;  DB 9;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 RIHWESASLL 15
           ||||||||||
Db       1 RIHWESASLL 10


RESULT 8
US-09-867-550-910
; Sequence 910, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 910
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)
;   OTHER INFORMATION: wherein Xaa may be any one of Arg or Cys or Gly or Ser
US-09-867-550-910

  Query Match             53.2%;  Score 42;  DB 10;  Length 91;
  Best Local Similarity   50.0%;  Pred. No. 4;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASL 14
           ||:  | ||:|  |
Db      72 SKVCSRFHWDSGVL 85


RESULT 9
US-09-764-864-816
; Sequence 816, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 816
;   LENGTH: 74
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (23)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-816

  Query Match             50.6%;  Score 40;  DB 10;  Length 74;
  Best Local Similarity   57.1%;  Pred. No. 6.7;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASL 14
           :|| | :| :||||
Db      36 TKIKHFLHQQSASL 49


RESULT 10
US-09-867-550-908
; Sequence 908, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 908
;   LENGTH: 64
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-867-550-908

  Query Match             49.4%;  Score 39;  DB 10;  Length 64;
  Best Local Similarity   46.2%;  Pred. No. 8.4;
  Matches    6;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESAS 13
           :::|||:|  ||:
Db      19 NRLTHRVHSASAN 31


RESULT 11
US-09-808-602-68
; Sequence 68, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
;  APPLICANT: Vernet, Corine A
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Herrman, John L
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Mezes, Peter S
;  APPLICANT:  MacDougall, John
;  TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
;  FILE REFERENCE: 15966-697 CIP
;  CURRENT APPLICATION NUMBER: US/09/808,602
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: 09/800,198
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: 60/186,596
;  PRIOR FILING DATE: 2000-03-03
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 68
;   LENGTH: 2012
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-808-602-68

  Query Match             48.1%;  Score 38;  DB 9;  Length 2012;
  Best Local Similarity   45.5%;  Pred. No. 3.9e+02;
  Matches    5;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy      3 ITHRIHWESAS 13
          :|| :|::| |
Db   1702 VTHTVHYQSVS 1712


RESULT 12
US-09-864-761-35988
; Sequence 35988, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35988
;   LENGTH: 40
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL079338.9
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.99
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
;   OTHER INFORMATION: EST_HUMAN HIT: BF541030.1, EVALUE 7.00e-17
;   OTHER INFORMATION: SWISSPROT HIT: P14336, EVALUE 3.60e+00

US-09-864-761-35988

  Query Match             46.8%;  Score 37;  DB 10;  Length 40;
  Best Local Similarity   45.5%;  Pred. No. 11;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
          | :|||  |::
Db      7 HLLHWEMKSVI 17


RESULT 13
US-09-864-761-44965
; Sequence 44965, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 44965
;   LENGTH: 69
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC016498.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.78
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
```

```
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.83
;   OTHER INFORMATION: EST_HUMAN HIT: BE877915.1, EVALUE 1.10e-02
US-09-864-761-44965

  Query Match           46.8%;  Score 37;  DB 10;  Length 69;
  Best Local Similarity  75.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       8 HWESASLL 15
           ||: ||||
Db      25 HWQGASLL 32


RESULT 14
US-09-864-761-34487
; Sequence 34487, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34487
;   LENGTH: 103
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006548.19
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.3
;   OTHER INFORMATION: SWISSPROT HIT: P41064, EVALUE 1.30e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AU140898.1, EVALUE 9.00e-26
US-09-864-761-34487

  Query Match           46.8%;  Score 37;  DB 10;  Length 103;
  Best Local Similarity  70.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 THRIHWESAS 13
           |||| |  ||
Db      58 THRIPWSLAS 67


RESULT 15
US-09-864-761-46114
; Sequence 46114, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
```

```
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46114
;   LENGTH: 134
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006548.20
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
;   OTHER INFORMATION: EST_HUMAN HIT: AU140898.1, EVALUE 6.00e-37
;   OTHER INFORMATION: SWISSPROT HIT: Q09312, EVALUE 2.00e+00
US-09-864-761-46114

  Query Match             46.8%;  Score 37;  DB 10;  Length 134;
  Best Local Similarity   70.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       4 THRIHWESAS 13
           |||| |  ||
Db      63 THRIPWSLAS 72

Search completed: February 21, 2003, 14:18:28
Job time : 14 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:11 ; Search time 48 Seconds
(without alignments)
30.042 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	100.0	1663	1	C3HU	complement C3 prec
2	52	65.8	267	2	A82997	hypothetical prote
3	52	65.8	726	2	A27602	complement C3 - ra
4	45	57.0	211	2	H83239	pseudouridine synt
5	44	55.7	516	2	S67037	SMP3 protein - yea
6	42	53.2	401	2	E82521	hypothetical prote
7	41	51.9	226	1	JQ0393	nodulation protein
8	41	51.9	229	2	AI3289	hypothetical cytos
9	41	51.9	336	2	F75508	mrr restriction sy
10	41	51.9	615	2	B86713	hypothetical prote
11	40.5	51.3	1417	2	H90670	probable invasin [
12	40.5	51.3	1417	2	D85521	probable adhesin e
13	40	50.6	259	2	T29569	hypothetical prote
14	40	50.6	343	2	T42129	probable acyltrans
15	39	49.4	228	2	AI2913	conserved hypothet
16	39	49.4	266	2	D97688	hypothetical prote
17	39	49.4	354	2	D41080	probable aldolase
18	39	49.4	406	2	T50894	hydroxyneurosporen
19	39	49.4	567	2	C69611	ABC transporter re
20	39	49.4	574	2	AB1790	ABC transporter re
21	39	49.4	574	2	AC1414	ABC transporter re
22	39	49.4	851	1	WMBEU9	gene UL9 protein -
23	38	48.1	151	2	S48796	transforming prote
24	38	48.1	242	2	C70895	hypothetical prote
25	38	48.1	248	2	AH0011	ferredoxin-NADP re
26	38	48.1	280	2	S46699	hypothetical prote
27	38	48.1	280	2	T21876	hypothetical prote
28	38	48.1	280	2	C86317	protein T10O22.23
29	38	48.1	300	2	G70943	hypothetical prote
30	38	48.1	398	2	A83277	hypothetical prote
31	38	48.1	399	2	AD0498	conserved hypothet
32	38	48.1	459	2	B82416	hypothetical prote
33	38	48.1	474	2	G75580	conserved hypothet
34	38	48.1	609	2	G84832	ATP-dependent RNA
35	38	48.1	778	2	T16111	hypothetical prote
36	38	48.1	1456	2	G86466	hypothetical prote
37	38	48.1	1896	2	T08851	Down syndrome cell
38	37.5	47.5	609	2	AB0500	glutamine-fructose
39	37	46.8	148	2	A86878	non-heme iron-bind
40	37	46.8	178	2	AF1368	B. subtilis YtmI p
41	37	46.8	178	2	AG1737	B. subtilis YtmI p
42	37	46.8	220	2	S16511	probable PRT1 prot
43	37	46.8	227	2	C82625	conserved hypothet
44	37	46.8	249	2	T17134	hypothetical prote
45	37	46.8	314	2	F70505	probable trna delt

ALIGNMENTS

RESULT 1
C3HU
complement C3 precursor [validated] - human
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subuni
C;Species: Homo sapiens (man)
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Dec-2000
C;Accession: A94065; A37999; A92187; A27603; A23435; A45830; B45830; A01257; A01258
R;de Bruijn, M.H.L.; Fey, G.H.
Proc. Natl. Acad. Sci. U.S.A. 82, 708-712, 1985
A;Title: Human complement component C3: cDNA coding sequence and derived primary stru
A;Reference number: A94065; MUID:85140166; PMID:2579379
A;Accession: A94065
A;Molecule type: mRNA
A;Residues: 1-1663 <DEB>
A;Cross-references: GB:K02765; NID:g179664; PIDN:AAA85332.1; PID:g179665
R;Vik, D.P.; Amiguet, P.; Moffat, G.J.; Fey, M.; Amiguet-Barras, F.; Wetsel, R.A.; T
Biochemistry 30, 1080-1085, 1991
A;Title: Structural features of the human C3 gene: intron/exon organization, transcri
A;Reference number: A37999; MUID:91113687; PMID:1703437
A;Contents: intron/exon structure of gene
A;Accession: A37999
A;Molecule type: DNA
A;Residues: 1-25 <VIK>
A;Cross-references: GB:M63423
A;Note: the authors translated the codon GGT for residue 6 as Leu, CCC for residue 7
R;Hugli, T.E.
J. Biol. Chem. 250, 8293-8301, 1975
A;Title: Human anaphylatoxin (C3a) from the third component of complement.
A;Reference number: A92187; MUID:76069169; PMID:1238393
A;Accession: A92187
A;Molecule type: protein
A;Residues: 672-680,'N',682-699,'Q',701-748 <HUG>
R;Daoudaki, M.E.; Becherer, J.D.; Lambris, J.D.
J. Immunol. 140, 1577-1580, 1988
A;Title: A 34-amino acid peptide of the third component of complement mediates prope
A;Reference number: A27603; MUID:88154452; PMID:3279119
A;Accession: A27603
A;Molecule type: protein
A;Residues: 1409-1563 <DAO>
R;Hellman, U.; Eggertsen, G.; Engstrom, A.; Sjoquist, J.
Biochem. J. 230, 353-361, 1985
A;Title: Amino acid sequence of the trypsin-generated C3d fragment from human comple
A;Reference number: A23435; MUID:86025442; PMID:3876831
A;Accession: A23435
A;Molecule type: protein
A;Residues: 1002-1012,'E',1014-1303 <HEL>
A;Note: sequence corresponding to residues 1072-1100 was not determined but was take
R;Poznansky, M.C.; Clissold, P.M.; Lachmann, P.J.
J. Immunol. 143, 1254-1258, 1989
A;Title: The difference between human C3F and C3S results from a single amino acid ch
3.
A;Reference number: A45830; MUID:89309808; PMID:2473125

A;Accession: A45830
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1212-1215,'N',1217-1223 <POZ>
A;Note: this is the C3S allele
A;Accession: B45830
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1212-1223 <PO2>
R;Dolmer, K.; Sottrup-Jensen, L.
FEBS Lett. 315, 85-90, 1993
A;Title: Disulfide bridges in human complement component C3b.
A;Reference number: S27041; MUID:93106233; PMID:8416818
A;Contents: annotation; disulfide bonds
C;Comment: The sequence shown is the C3 fast (C3F) allele, which is found mainly in Cauc
C;Comment: Complement C3 contains two chains, formed by removal of four residues and lin
alternative complement pathways, releases the C3a anaphylatoxin from the amino end of t
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign pa
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by prc
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Genetics:
A;Gene: GDB:C3
A;Cross-references: GDB:119044; OMIM:120700
A;Map position: 19p13.3-19p13.3
A;Note: contains 41 exons
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprotein;
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-667/Product: complement C3 and C3b beta chain #status predicted <C3BB>
F;23-667,672-1663/Product: complement C3 #status predicted <CC3>
F;23-667,749-1663/Product: C3b #status predicted <C3B>
F;672-1663/Product: complement C3 alpha chain #status predicted <CC3A>
F;672-748/Product: C3a anaphylatoxin #status predicted <C3T>
F;749-1663/Product: C3b alpha' chain #status predicted <C3BA>
F;946-1303/Product: C3dk fragment #status predicted <CDK>
F;955-1303/Product: C3dg fragment #status predicted <CDG>
F;955-1001/Product: C3g fragment #status predicted <C3G>
F;1002-1303/Product: C3d fragment #status experimental <C3D>
F;1424-1457/Region: properdin binding
F;85,939/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;559-816,627-662,693-720,694-727,707-728,873-1513,1101-1158,1358-1489,1389-1458,1506-15
F;748-749/Cleavage site: Arg-Ser (C3 convertase) #status predicted
F;954-955/Cleavage site: Arg-Glu (complement factor I) #status predicted
F;1010-1013/Cross-link: thiolester (Cys-Gln) #status experimental
F;1303-1304/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1320-1321/Cleavage site: Arg-Ser (complement factor I) #status predicted
F;1617/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match 100.0%; Score 79; DB 1; Length 1663;
  Best Local Similarity 100.0%; Pred. No. 2.6e-05;
  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SKITHRIHWESASLL 15
||||||||||||||||
Db 1305 SKITHRIHWESASLL 1319

RESULT 2
A82997
hypothetical protein PA5194 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A82997
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A82997
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <STO>
A;Cross-references: GB:AE004932; GB:AE004091; NID:g9951493; PIDN:AAG08579.1; GSPDB:G
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA5194

  Query Match 65.8%; Score 52; DB 2; Length 267;
  Best Local Similarity 57.1%; Pred. No. 0.17;
  Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SKITHRIHWESASL 14
:|| | :||: |||
Db 118 AKIAHHLHWQHASL 131

RESULT 3
A27602
complement C3 - rabbit (fragment)
N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) C3b subun
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 15-Dec-1988 #sequence_revision 07-Oct-1994 #text_change 16-Jul-1999
C;Accession: A27602
R;Kusano, M.; Choi, N.H.; Tomita, M.; Yamamoto, K.; Migita, S.; Sekiya, T.; Nishimura
Immunol. Invest. 15, 365-378, 1986
A;Title: Nucleotide sequence of cDNA and derived amino acid sequence of rabbit comple
A;Reference number: A27602; MUID:87006907; PMID:3019881
A;Accession: A27602
A;Molecule type: mRNA
A;Residues: 1-726 <KUS>
A;Cross-references: GB:M32434; NID:g164862; PIDN:AAA31190.1; PID:g164863
C;Comment: Complement C3 contains two chains, formed by removal of four residues and
alternative complement pathways, releases the C3a anaphylatoxin from the amino end o
rnative-complement-pathway C3/C5 convertase.
C;Comment: C3a anaphylatoxin is a vasoactive peptide and a mediator of inflammation.
C;Comment: C3b, with its highly reactive thiol group, binds to the surface of foreign
e classical-complement-pathway C3/C5 convertase. The activity of C3b is regulated by
C;Comment: The major site of synthesis of this plasma protein is the liver.
C;Superfamily: alpha-2-macroglobulin
C;Keywords: acute phase; complement alternate pathway; complement pathway; glycoprote

  Query Match 65.8%; Score 52; DB 2; Length 726;
  Best Local Similarity 66.7%; Pred. No. 0.52;
  Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 SKITHRIHWESASLL 15
| : ||| |:|||||
Db 368 SPVKHRIVWDSASLL 382

RESULT 4
H83239
pseudouridine synthase RluA PA3246 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83239
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83239
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-211 <STO>
A;Cross-references: GB:AE004747; GB:AE004091; NID:g9949362; PIDN:AAG06634.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: rluA; PA3246

  Query Match 57.0%; Score 45; DB 2; Length 211;
  Best Local Similarity 40.0%; Pred. No. 2.2;

```
  Matches    6;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          ::| ||: ||:: |:
Db     50 ARIVHRLDWETSGLM 64


RESULT 5
S67037
SMP3 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O3527; protein YOR149c
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 21-Jul-2000
C;Accession: S67037; S13750
R;Bordonne, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarassov, I.A.; Winsc
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67032
A;Accession: S67037
A;Molecule type: DNA
A;Residues: 1-516 <BOR>
A;Cross-references: EMBL:Z75057; NID:g1420374; PID:e252038; PID:g1420375; MIPS:YOR149c
A;Experimental source: strain S288C
R;Irie, K.; Araki, H.; Oshima, Y.
Mol. Gen. Genet. 225, 257-265, 1991
A;Title: Mutations in a Saccharomyces cerevisiae host showing increased holding stabilit
A;Reference number: S13750; MUID:91172125; PMID:2005867
A;Accession: S13750
A;Molecule type: DNA
A;Residues: 1-121,'IK',124-162,'G',164-168,'R',170-278,'L',280-516 <IRI>
A;Cross-references: EMBL:X58121; NID:g4497; PIDN:CAA41123.1; PID:g4498
C;Genetics:
A;Gene: SGD:SMP3
A;Cross-references: SGD:S0005675; MIPS:YOR149c
A;Map position: 15R
C;Keywords: transmembrane protein
F;9-25/Domain: transmembrane #status predicted <TM1>
F;189-205/Domain: transmembrane #status predicted <TM2>
F;215-231/Domain: transmembrane #status predicted <TM3>
F;271-287/Domain: transmembrane #status predicted <TM4>
F;344-360/Domain: transmembrane #status predicted <TM5>

  Query Match             55.7%;  Score 44;  DB 2;  Length 516;
  Best Local Similarity   63.6%;  Pred. No. 9;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
          :|:||:| |||
Db    207 YRVHWKSFSLL 217


RESULT 6
E82521
hypothetical protein XF2735 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: E82521
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-401 <SIM>
A;Cross-references: GB:AE004080; GB:AE003849; NID:g9107971; PIDN:AAF85520.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laign
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martin
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2735

  Query Match             53.2%;  Score 42;  DB 2;  Length 401;
  Best Local Similarity   45.5%;  Pred. No. 15;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 ITHRIHWESAS 13
          : ||:||:  |
Db    334 LAHRVHWDEES 344


RESULT 7
JQ0393
nodulation protein nodA - Azorhizobium caulinodans
N;Alternate names: hypothetical 24.9K protein
C;Species: Azorhizobium caulinodans
A;Note: host Sesbania rostrata
C;Date: 07-Sep-1990 #sequence_revision 27-Jan-1995 #text_change 16-Jul-1999
C;Accession: JQ0393
R;Goethals, K.; Gao, M.; Tomekpe, K.; Van Montagu, M.; Holsters, M.
Mol. Gen. Genet. 219, 289-298, 1989
A;Title: Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans: nucleotide
A;Reference number: JQ0393; MUID:90136519; PMID:2615763
A;Accession: JQ0393
A;Molecule type: DNA
A;Residues: 1-226 <GOE>
A;Cross-references: GB:L18897; NID:g1293899; PIDN:AAB51162.1; PID:g310292
A;Experimental source: strain ORS571
C;Comment: This is one of the proteins, coded by nodulation genes, that are required
C;Genetics:
A;Gene: nodA
C;Superfamily: nodulation protein nodA
C;Keywords: nodulation

  Query Match             51.9%;  Score 41;  DB 1;  Length 226;
  Best Local Similarity   63.6%;  Pred. No. 12;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWES 11
          ||:| |: |||
Db     33 SKVTWRVAWES 43


RESULT 8
AI3289
hypothetical cytosolic protein BMEI0303 [imported] - Brucella melitensis (strain 16M
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AI3289
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivano
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Le
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella meli
A;Reference number: AD3252; PMID:11756688
A;Accession: AI3289
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51484.1; PID:g17982196; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0303
A;Map position: I
C;Superfamily: Rickettsia prowazekii hypothetical protein RP073
```

```
  Query Match             51.9%;  Score 41;  DB 2;  Length 229;
  Best Local Similarity   53.8%;  Pred. No. 12;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 KITHRIHWESASL 14
           :| :| || ||:|
Db     136 QIRNRTHWNSANL 148


RESULT 9
F75508
mrr restriction system protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75508
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75508
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <WHI>
A;Cross-references: GB:AE001910; GB:AE000513; NID:g6458198; PIDN:AAF10088.1; PID:g645819
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0508
A;Map position: 1

  Query Match             51.9%;  Score 41;  DB 2;  Length 336;
  Best Local Similarity   50.0%;  Pred. No. 19;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASL 14
           ||: ||| |  ::|
Db      72 SKVRHRIAWACSNL 85


RESULT 10
B86713
hypothetical protein cydC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 24-Aug-2001
C;Accession: B86713
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: B86713
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-615 <STO>
A;Cross-references: GB:AE005176; PID:g12723617; PIDN:AAK04804.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: cydC
C;Superfamily: Mycobacterium tuberculosis probable ABC transporter cydD; ATP-binding cas

  Query Match             51.9%;  Score 41;  DB 2;  Length 615;
  Best Local Similarity   66.7%;  Pred. No. 36;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 THRIHWESA 12
           |||:|| |:
Db     509 THRLHWLSS 517


RESULT 11
H90670
probable invasin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H90670
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1417 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33759.1; PID:g13359793; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0336

  Query Match             51.3%;  Score 40.5;  DB 2;  Length 1417;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    3;  Gaps    1;

Qy       1 SKITH---RIHWESASLL 15
           || ||    : ||: |||
Db     385 SKATHGLKNVQWEAPSLL 402


RESULT 12
D85521
probable adhesin eaeH [imported] - Escherichia coli (strain O157:H7, substrain EDL933
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85521
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1417 <STO>
A;Cross-references: GB:AE005174; NID:g12513096; PIDN:AAG54632.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: eaeH

  Query Match             51.3%;  Score 40.5;  DB 2;  Length 1417;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    3;  Gaps    1;

Qy       1 SKITH---RIHWESASLL 15
           || ||    : ||: |||
Db     385 SKATHGLKNVQWEAPSLL 402


RESULT 13
T29569
hypothetical protein C44C1.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T29569
R;Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C44C1.
A;Reference number: Z20642
A;Accession: T29569
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-259 <BRA>
A;Cross-references: EMBL:U41030; PIDN:AAA82366.1; CESP:C44C1.1
C;Genetics:
A;Gene: CESP:C44C1.1
A;Introns: 34/3; 82/1; 105/1; 146/2
```

```
C;Superfamily: Caenorhabditis elegans hypothetical protein C44C1.1

  Query Match             50.6%;  Score 40;  DB 2;  Length 259;
  Best Local Similarity   75.0%;  Pred. No. 21;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 THRIHWES 11
           || :|||| 
Db     185 THVLHWES 192


RESULT 14
T42129
probable acyltransferase (EC 2.3.1.-) - Escherichia coli plasmid pO157
C;Species: Escherichia coli
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C;Accession: T42129; T00321
R;Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A;Title: The complete DNA sequence and analysis of the large virulence plasmid of Escher
A;Reference number: Z22068; MUID:98391744; PMID:9722640
A;Accession: T42129
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-343 <BUR>
A;Cross-references: EMBL:AF074613; PIDN:AAC70097.1
A;Experimental source: strain EDL933; serotype O157:H7
R;Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota,
 S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A;Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrha
A;Reference number: Z14127; MUID:98290540; PMID:9628576
A;Accession: T00321
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 12-343 <MAK>
A;Cross-references: EMBL:AB011549; NID:g4589740; PIDN:BAA31840.1; PID:g3337081
A;Experimental source: strain EHEC O157:H7, substrain RIMD 0509952
C;Genetics:
A;Genome: plasmid pO157
A;Note: L7029
C;Keywords: acyltransferase

  Query Match             50.6%;  Score 40;  DB 2;  Length 343;
  Best Local Similarity   42.9%;  Pred. No. 29;
  Matches    6;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 KITHRIHWESASLL 15
           ||:||| |    ::
Db     140 KISHRIRWNGLEIV 153


RESULT 15
AI2913
conserved hypothetical protein Atu2746 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002
C;Accession: AI2913
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AI2913
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-228 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43727.1; PID:g17741259; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2746
A;Map position: circular chromosome
C;Superfamily: Rickettsia prowazekii hypothetical protein RP073

  Query Match             49.4%;  Score 39;  DB 2;  Length 228;
  Best Local Similarity   53.8%;  Pred. No. 27;
  Matches    7;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 KITHRIHWESASL 14
           :|  | || ||:|
Db     130 QIRDRTHWNSANL 142


Search completed: February 21, 2003, 14:17:35
Job time : 53 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:11 ; Search time 11 Seconds
(without alignments)
56.559 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	100.0	1663	1	CO3_HUMAN	P01024 homo sapien
2	52	65.8	726	1	CO3_RABIT	P12247 oryctolagus
3	44	55.7	516	1	SMP3_YEAST	Q04174 saccharomyc
4	41	51.9	226	1	NODA_AZOCA	Q07739 azorhizobiu
5	40	50.6	396	1	RT09_HUMAN	P82933 homo sapien
6	39	49.4	336	1	PTXD_PSEST	O69054 pseudomonas
7	39	49.4	354	1	ALF2_RHOSH	P29271 rhodobacter
8	39	49.4	567	1	CYDC_BACSU	P94366 bacillus su
9	39	49.4	851	1	OBP_HSV11	P10193 herpes simp
10	38	48.1	242	1	YA85_MYCTU	O53433 mycobacteri
11	38	48.1	280	1	GEM2_HUMAN	O14893 homo sapien
12	38	48.1	280	1	YHM7_YEAST	P38790 saccharomyc
13	38	48.1	2012	1	DSCA_HUMAN	O60469 homo sapien
14	37.5	47.5	608	1	GLMS_YERPE	Q8z9s8 y glucosami
15	37	46.8	220	1	PRT1_PICAN	P12806 pichia angu
16	37	46.8	269	1	GEM2_RAT	Q9qzp1 rattus norv
17	37	46.8	314	1	MIAA_MYCTU	O33232 mycobacteri
18	37	46.8	345	1	SEP3_HUMAN	Q9uh03 homo sapien
19	37	46.8	465	1	SEP3_MOUSE	Q9z1s5 mus musculu
20	37	46.8	587	1	T9S3_MOUSE	Q9et30 mus musculu
21	37	46.8	589	1	T9S3_HUMAN	Q9hd45 homo sapien
22	37	46.8	698	1	TNPX_ECOLI	Q00042 escherichia
23	37	46.8	1663	1	CO3_RAT	P01026 rattus norv
24	37	46.8	4385	1	YP73_CAEEL	Q09222 caenorhabdi
25	36.5	46.2	847	1	VAV3_MOUSE	Q9r0c8 mus musculu
26	36	45.6	150	1	VG50_BPT4	P15075 bacteriopha
27	36	45.6	166	1	LITA_HUMAN	P05451 homo sapien
28	36	45.6	166	1	LITB_HUMAN	P48304 homo sapien
29	36	45.6	173	1	LIT2_MOUSE	Q08731 mus musculu
30	36	45.6	175	1	RMP2_HUMAN	O60895 homo sapien
31	36	45.6	260	1	CAH1_MOUSE	P13634 mus musculu
32	36	45.6	260	1	CAH1_SHEEP	P48282 ovis aries
33	36	45.6	320	1	NODI_AZOCA	Q07756 azorhizobiu
34	36	45.6	335	1	VMSA_HPBHE	P13847 heron hepat
35	36	45.6	508	1	YM05_ARCFU	O28078 archaeoglob
36	36	45.6	518	1	GGT_SYNY3	P74181 synechocyst
37	36	45.6	569	1	U171_HUMAN	Q12980 homo sapien
38	36	45.6	608	1	GLMS_BUCAI	P57138 b glucosami
39	36	45.6	625	1	T9S4_HUMAN	Q92544 homo sapien
40	36	45.6	751	1	TREA_YEAST	P32356 saccharomyc
41	36	45.6	843	1	VAV_RAT	P54100 rattus norv
42	36	45.6	845	1	VAV_MOUSE	P27870 mus musculu
43	36	45.6	847	1	VAV3_HUMAN	Q9ukw4 homo sapien
44	36	45.6	944	1	VP35_YEAST	P34110 saccharomyc
45	36	45.6	1157	1	YE56_CAEEL	P90747 caenorhabdi

ALIGNMENTS

RESULT 1
CO3_HUMAN
ID CO3_HUMAN STANDARD; PRT; 1663 AA.
AC P01024;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement C3 precursor [Contains: C3a anaphylatoxin].
GN C3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85140166; PubMed=2579379;
RA de Bruijn M.H.L., Fey G.H.;
RT "Human complement component C3: cDNA coding sequence and derived
RT primary structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:708-712(1985).
RN [2]
RP SEQUENCE OF 672-748.
RX MEDLINE=76069169; PubMed=1238393;
RA Hugli T.E.;
RT "Human anaphylatoxin (C3a) from the third component of complement.
RT Primary structure.";
RL J. Biol. Chem. 250:8293-8301(1975).
RN [3]
RP SEQUENCE OF 955-966, AND SUBUNITS.
RC TISSUE=Serum;
RX MEDLINE=95293954; PubMed=7539791;
RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,
RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;
RT "Identification of angiotensinogen and complement C3dg as novel
RT proteins binding the proform of eosinophil major basic protein in
RT human pregnancy serum and plasma.";
RL J. Biol. Chem. 270:13645-13651(1995).
RN [4]
RP SEQUENCE OF 988-1036.
RX MEDLINE=82174534; PubMed=6175959;
RA Thomas M.L., Janatova J., Gray W.R., Tack B.F.;
RT "Third component of human complement: localization of the internal
RT thiolester bond.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1054-1058(1982).
RN [5]
RP SEQUENCE OF 1409-1563.
RX MEDLINE=88154452; PubMed=3279119;
RA Daoudaki M.E., Becherer J.D., Lambris J.D.;
RT "A 34-amino acid peptide of the third component of complement
RT mediates properdin binding.";
RL J. Immunol. 140:1577-1580(1988).
RN [6]
RP STRUCTURE BY NMR OF C3A.
RX MEDLINE=88276894; PubMed=3260670;
RA Nettesheim D.G., Edalji R.P., Mollison K.W., Greer J.,
RA Zuiderweg E.R.P.;

```
RT   "Secondary structure of complement component C3a anaphylatoxin in
RT   solution as determined by NMR spectroscopy: differences between
RT   crystal and solution conformations.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:5036-5040(1988).
RN   [7]
RP   MUTAGENESIS OF THIOESTER BOND REGION.
RX   MEDLINE=92250565; PubMed=1577777;
RA   Isaac L., Isenman D.E.;
RT   "Structural requirements for thioester bond formation in human
RT   complement component C3. Reassessment of the role of thioester bond
RT   integrity on the conformation of C3.";
RL   J. Biol. Chem. 267:10062-10069(1992).
RN   [8]
RP   DISULFIDE BONDS.
RX   MEDLINE=93106233; PubMed=8416818;
RA   Dolmer K., Sottrup-Jensen L.;
RT   "Disulfide bridges in human complement component C3b.";
RL   FEBS Lett. 315:85-90(1993).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 996-1303.
RX   MEDLINE=98259089; PubMed=9596584;
RA   Nagar B., Jones R.G., Diefenbach R.J., Isenman D.E., Rini J.M.;
RT   "X-ray crystal structure of C3d: a C3 fragment and ligand for
RT   complement receptor 2.";
RL   Science 280:1277-1281(1998).
RN   [10]
RP   VARIANT C3F/S.
RX   MEDLINE=89309808; PubMed=2473125;
RA   Poznansky M.C., Clissold P.M., Lachmann P.J.;
RT   "The difference between human C3F and C3S results from a single amino
RT   acid change from an asparagine to an aspartate residue at position
RT   1216 on the alpha-chain of the complement component, C3.";
RL   J. Immunol. 143:1254-1258(1989).
RN   [11]
RP   ERRATUM (RETRACTION OF ABOVE ARTICLE).
RX   MEDLINE=90063087; PubMed=2584723;
RA   Poznansky M.C., Clissold P.M., Lachmann P.J.;
RL   J. Immunol. 143:3860-3862(1989).
RN   [12]
RP   VARIANTS GLY-102 AND PRO-314.
RX   MEDLINE=91011240; PubMed=1976733;
RA   Botto M., Yong Fong K., So A.K., Koch C., Walport M.J.;
RT   "Molecular basis of polymorphisms of human complement component C3.";
RL   J. Exp. Med. 172:1011-1017(1990).
RN   [13]
RP   VARIANT ASN-549.
RX   MEDLINE=95050640; PubMed=7961791;
RA   Singer L., Whitehead W.T., Akama H., Katz Y., Fishelson Z.,
RA   Wetsel R.A.;
RT   "Inherited human complement C3 deficiency. An amino acid substitution
RT   in the beta-chain (ASP549 to ASN) impairs C3 secretion.";
RL   J. Biol. Chem. 269:28494-28499(1994).
RN   [14]
RP   VARIANT GLN-1320.
RA   Watanabe Y., Matsui N., Yan K., Nishimukai H., Tokunaga K.,
RA   Juji T., Kobayashi N., Kohsaka T.;
RT   "A novel C3 allotype C3'F02'has an amino acid substitution that may
RT   inhibit iC3b synthesis and cause C3-hypocomplementemia.";
RL   Mol. Immunol. 30:62-62(1993).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,
CC       C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC       INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC       PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC       BASOPHILIC LEUKOCYTES.
CC   -!- SUBUNIT: C3 precursor is first processed by the removal of 4 Arg
CC       residues, forming two chains, beta and alpha, linked by a
CC       disulfide bond. C3 convertase activates C3 by cleaving the alpha
CC       chain, releasing C3A anaphylatoxin and generating C3B (beta chain
CC       + alpha' chain). During pregnancy, C3dg exists as a complex
CC       (probably a 2:2:2 heterohexamer) with AGT and the proform of PRG2.
CC   -!- POLYMORPHISM: THERE ARE TWO ALLELES: C3S (C3 SLOW), THE MOST
CC       COMMON ALLELE IN ALL RACES AND C3F (C3 FAST), RELATIVELY FREQUENT
CC       IN CAUCASOIDS, LESS COMMON IN BLACK AMERICAN, EXTREMELY RARE IN
CC       ORIENTALS.
CC   -!- DISEASE: C3 DEFICIENCY CAUSES A SUSCEPTIBILITY TO PYOGENIC
CC       INFECTION.
CC   -!- MISCELLANEOUS: C3B IS RAPIDLY SPLIT IN TWO POSITIONS BY FACTOR I
CC       AND A COFACTOR TO FORM IC3B (INACTIVATED C3B) AND C3F WHICH IS
CC       RELEASED.
CC   -!- MISCELLANEOUS: IC3B IS THE SLOWLY CLEAVED (POSSIBLY BY FACTOR I)
CC       TO FORM C3C AND C3DG. OTHER PROTEASES PRODUCE OTHER FRAGMENTS SUCH
CC       AS C3D OR C3G.
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02765; AAA85332.1; -.
DR   PIR; A01257; C3HU.
DR   PIR; A27603; A27603.
DR   PDB; 1C3D; 18-NOV-98.
DR   SWISS-2DPAGE; P01024; HUMAN.
DR   Siena-2DPAGE; P01024; -.
DR   Genew; HGNC:1318; C3.
DR   MIM; 120700; -.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein; Signal; Polymorphism;
KW   Disease mutation; 3D-structure.
FT   SIGNAL       1     22
FT   CHAIN       23   1663       COMPLEMENT C3.
FT   CHAIN       23    667       COMPLEMENT C3, BETA CHAIN.
FT   CHAIN      672   1663       COMPLEMENT C3, ALPHA CHAIN.
FT   PEPTIDE    672    748       C3A ANAPHYLATOXIN.
FT   CHAIN      749   1663       C3B ALPHA' CHAIN.
FT   PEPTIDE    749    954       C3C FRAGMENT.
FT   PEPTIDE    955   1303       C3DG FRAGMENT.
FT   PEPTIDE    955   1001       C3G FRAGMENT.
FT   PEPTIDE   1002   1303       C3D FRAGMENT.
FT   PEPTIDE   1304   1320       C3F FRAGMENT.
FT   SITE       748    749       CLEAVAGE (BY C3 CONVERTASE).
FT   SITE       954    955       CLEAVAGE (BY FACTOR I) (POTENTIAL).
FT   SITE      1303   1304       CLEAVAGE (BY FACTOR I).
FT   SITE      1320   1321       CLEAVAGE (BY FACTOR I).
FT   DOMAIN     693    728       ANAPHYLATOXIN-LIKE.
FT   DOMAIN    1424   1456       PROPERDIN-BINDING.
FT   DISULFID   559    816       INTERCHAIN.
FT   DISULFID   627    662
FT   DISULFID   693    720
FT   DISULFID   694    727
FT   DISULFID   707    728
```

```
  Query Match             100.0%;  Score 79;  DB 1;  Length 1663;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          |||||||||||||||
Db   1305 SKITHRIHWESASLL 1319


RESULT 2
CO3_RABIT
ID   CO3_RABIT      STANDARD;      PRT;   726 AA.
AC   P12247;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Complement C3 alpha chain (Fragment).
GN   C3.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87006907; PubMed=3019881;
RA   Kusano M., Choi N.H., Tomita M., Yamamoto K., Migita S., Sekiya T.,
RA   Nishimura S.;
RT   "Nucleotide sequence of cDNA and derived amino acid sequence of
RT   rabbit complement component C3 alpha-chain.";
RL   Immunol. Invest. 15:365-378(1986).
CC   -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC       COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL
CC       REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.
CC       AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE
CC       THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.
CC   -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG
CC       RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC       BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN,
CC       RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'
CC       CHAIN).
CC   -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M32434; AAA31190.1; -.
DR   PIR; A27602; A27602.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001599; MacroglobInA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01759; NTR; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; PARTIAL.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; PARTIAL.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
KW   Complement pathway; Complement alternate pathway; Plasma;
KW   Inflammatory response; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN        <1    726       COMPLEMENT C3 ALPHA CHAIN.
FT   THIOLEST     73     76
FT   CARBOHYD      2      2       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    680    680       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   726 AA;  81844 MW;  F4B4C35D461300E9 CRC64;

  Query Match             65.8%;  Score 52;  DB 1;  Length 726;
  Best Local Similarity   66.7%;  Pred. No. 0.19;
  Matches   10;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          | : ||| |:|||||
Db    368 SPVKHRIVWDSASLL 382


RESULT 3
SMP3_YEAST
ID   SMP3_YEAST     STANDARD;      PRT;   516 AA.
AC   Q04174; Q99400;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SMP3 protein.
GN   SMP3 OR YOR149C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NBW5;
RX   MEDLINE=91172125; PubMed=2005867;
RA   Irie K., Araki H., Oshima Y.;
RT   "Mutations in a Saccharomyces cerevisiae host showing increased
RT   holding stability of the heterologous plasmid pSR1.";
RL   Mol. Gen. Genet. 225:257-265(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1678;
RA   Ayadi A., Bordonne R., Camasses A., Madania A., Poch O.,
RA   Tarassov I.A., Winsor B., Martin R.P.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL PROTEIN INVOLVED IN PLASMID MAINTENANCE WITH
CC       SMP2.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO S.POMBE SPAC4G8.12C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58121; CAA41123.1; -.
DR   EMBL; U55020; AAC49635.1; -.
DR   EMBL; Z75057; CAA99355.1; -.
DR   PIR; S13750; S13750.
DR   SGD; S0005675; SMP3.
KW   Transmembrane.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM     61     81       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    211    231       POTENTIAL.
FT   TRANSMEM    271    291       POTENTIAL.
FT   TRANSMEM    296    316       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    349    369       POTENTIAL.
FT   CONFLICT    122    123       MQ -> IK (IN REF. 1).
FT   CONFLICT    163    163       E -> G (IN REF. 1).
FT   CONFLICT    169    169       S -> R (IN REF. 1).
FT   CONFLICT    279    279       V -> L (IN REF. 1).
SQ   SEQUENCE   516 AA;  59900 MW;  8D8404622CB69534 CRC64;

  Query Match             55.7%;  Score 44;  DB 1;  Length 516;
  Best Local Similarity   63.6%;  Pred. No. 3.3;
  Matches    7;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
```

```
          :|:||:| |||
Db    207 YRVHWKSFSLL 217


RESULT 4
NODA_AZOCA
ID   NODA_AZOCA     STANDARD;      PRT;   226 AA.
AC   Q07739;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nodulation protein A (EC 2.3.1.-).
GN   NODA.
OS   Azorhizobium caulinodans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Hyphomicrobium group; Azorhizobium.
OX   NCBI_TaxID=7;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ORS571;
RX   MEDLINE=90136519; PubMed=2615763;
RA   Goethals K., Gao M., Tomekpe K., van Montagu M., Holsters M.;
RT   "Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans:
RT   nucleotide sequence and plant-inducible expression.";
RL   Mol. Gen. Genet. 219:289-298(1989).
CC   -!- FUNCTION: N-ACYLTRANSFERASE REQUIRED FOR NODULATION. ACTS IN THE
CC       PRODUCTION OF A SMALL, HEAT-STABLE COMPOUND (NOD) THAT STIMULATES
CC       MITOSIS IN VARIOUS PLANT PROTOPLASTS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NODA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L18897; AAB51162.1; -.
DR   PIR; JQ0393; JQ0393.
DR   InterPro; IPR003484; NodA.
DR   Pfam; PF02474; NodA; 1.
DR   PROSITE; PS01349; NODA; 1.
KW   Transferase; Acyltransferase; Nodulation.
SQ   SEQUENCE   226 AA;  24915 MW;  F1992B421A002315 CRC64;

  Query Match             51.9%;  Score 41;  DB 1;  Length 226;
  Best Local Similarity   63.6%;  Pred. No. 4.6;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWES 11
          ||:| |: |||
Db     33 SKVTWRVAWES 43


RESULT 5
RT09_HUMAN
ID   RT09_HUMAN     STANDARD;      PRT;   396 AA.
AC   P82933;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   28S ribosomal protein S9, mitochondrial precursor (MRP-S9).
GN   MRPS9 OR RPMS9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.

RN   [2]
RP   IDENTIFICATION.
RX   MEDLINE=21276436; PubMed=11279123;
RA   Koc E.C., Burkhart W., Blackburn K., Moseley A., Spremulli L.L.;
RT   "The small subunit of the mammalian mitochondrial ribosome:
RT   identification of the full complement of ribosomal proteins present.";
RL   J. Biol. Chem. 276:19363-19374(2001).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BF034318; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR000754; Ribosomal_S9.
DR   Pfam; PF00380; Ribosomal_S9; 1.
DR   ProDom; PD001627; Ribosomal_S9; 1.
DR   PROSITE; PS00360; RIBOSOMAL_S9; 1.
KW   Ribosomal protein; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    396       28S RIBOSOMAL PROTEIN S9.
SQ   SEQUENCE   396 AA;  45822 MW;  A4ECC6FD3F7FE9AE CRC64;

  Query Match             50.6%;  Score 40;  DB 1;  Length 396;
  Best Local Similarity   54.5%;  Pred. No. 12;
  Matches    6;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
          |: ||:: |||
Db    175 HQSHWQAKSLL 185


RESULT 6
PTXD_PSEST
ID   PTXD_PSEST     STANDARD;      PRT;   336 AA.
AC   O69054;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phosphonate dehydrogenase (EC 1.20.1.1) (NAD-dependent phosphite
DE   dehydrogenase).
GN   PTXD.
OS   Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=316;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WM88;
RX   MEDLINE=99008986; PubMed=9791102;
RA   Metcalf W.W., Wolfe R.S.;
RT   "Molecular genetic analysis of phosphite and hypophosphite oxidation
RT   by Pseudomonas stutzeri WM88.";
RL   J. Bacteriol. 180:5547-5558(1998).
RN   [2]
RP   SEQUENCE OF 1-15, FUNCTION, ACTIVITY, COFACTOR, ENZYME REGULATION,
RP   SUBUNIT, INDUCTION, AND MASS SPECTROMETRY.
RC   STRAIN=WM88;
RX   MEDLINE=21264507; PubMed=11278981;
RA   Costas A.M.G., White A.K., Metcalf W.W.;
RT   "Purification and characterization of a novel phosphorus-oxidizing
RT   enzyme from Pseudomonas stutzeri WM88.";
RL   J. Biol. Chem. 276:17429-17436(2001).
CC   -!- FUNCTION: Catalyzes phosphite (phosphonate) oxidation.
CC   -!- CATALYTIC ACTIVITY: Phosphonate + NAD(+) + H(2)O = phosphate +
CC       NADH.
CC   -!- ENZYME REGULATION: Inhibited by NaCl, NADH and sulfite.
CC   -!- SUBUNIT: Homodimer.
```

```
CC   -!- INDUCTION: By phosphate starvation.
CC   -!- MASS SPECTROMETRY: MW=36413; MW_ERR=18; METHOD=MALDI.
CC   -!- MISCELLANEOUS: Its optimum pH is between 7.25 and 7.75 and optimum
CC       temperature is 35 degrees Celsius.
CC   -!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC       DEHYDROGENASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF061070; AAC71709.1; -.
DR   HSSP; P36234; 1GDH.
DR   InterPro; IPR002162; D_2hydac_dh.
DR   Pfam; PF00389; 2-Hacid_DH; 1.
DR   Pfam; PF02826; 2-Hacid_DH_C; 1.
DR   PROSITE; PS00065; D_2_HYDROXYACID_DH_1; FALSE_NEG.
DR   PROSITE; PS00670; D_2_HYDROXYACID_DH_2; FALSE_NEG.
DR   PROSITE; PS00671; D_2_HYDROXYACID_DH_3; FALSE_NEG.
KW   Oxidoreductase; NAD.
FT   ACT_SITE    237    237       SUBSTRATE-BINDING (BY SIMILARITY).
FT   ACT_SITE    266    266       BY SIMILARITY.
FT   ACT_SITE    292    292       BY SIMILARITY.
SQ   SEQUENCE   336 AA;  36415 MW;  7F55D246CA4454F7 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 336;
  Best Local Similarity   61.5%;  Pred. No. 15;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 ITHRIHWESASLL 15
          ||||:| |   ||
Db      7 ITHRVHDEILQLL 19


RESULT 7
ALF2_RHOSH
ID   ALF2_RHOSH     STANDARD;      PRT;   354 AA.
AC   P29271;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Fructose-bisphosphate aldolase II (EC 4.1.2.13).
GN   CFXB.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92041881; PubMed=1939098;
RA   Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;
RT   "Identification, expression, and deduced primary structure of
RT   transketolase and other enzymes encoded within the form II CO2
RT   fixation operon of Rhodobacter sphaeroides.";
RL   J. Biol. Chem. 266:20447-20452(1991).
CC   -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC       phosphate + D-glyceraldehyde 3-phosphate.
CC   -!- COFACTOR: ZINC.
CC   -!- PATHWAY: Glycolysis; sixth step.
CC   -!- PATHWAY: PART OF REDUCTIVE PENTOSE PHOSPHATE PATHWAY OR CALVIN
CC       CYCLE OF PHOTOSYNTHETIC CARBON DIOXIDE ASSIMILATION.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II
CC       RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.
CC   -!- SIMILARITY: BELONGS TO CLASS II FRUCTOSE-BISPHOSPHATE ALDOLASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M68914; AAA26157.1; -.
DR   PIR; D41080; D41080.
DR   InterPro; IPR000771; F_bP_aldolase.
DR   Pfam; PF01116; F_bP_aldolase; 1.
DR   ProDom; PD002376; F_bP_aldolase; 1.
DR   TIGRFAMs; TIGR00167; cbbA; 1.
DR   PROSITE; PS00602; ALDOLASE_CLASS_II_1; 1.
DR   PROSITE; PS00806; ALDOLASE_CLASS_II_2; 1.
KW   Lyase; Glycolysis; Zinc; Calvin cycle; Multigene family.
FT   METAL       81     81       ZINC (BY SIMILARITY).
FT   METAL       84     84       ZINC (BY SIMILARITY).
SQ   SEQUENCE   354 AA;  38269 MW;  9F547E84FC72ACF5 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 354;
  Best Local Similarity   35.7%;  Pred. No. 16;
  Matches    5;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASL 14
          ::::|  ||  ||:
Db    126 ARVSHMAHWVGASV 139


RESULT 8
CYDC_BACSU
ID   CYDC_BACSU     STANDARD;      PRT;   567 AA.
AC   P94366;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transport ATP-binding protein cydC.
GN   CYDC.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE=97124196; PubMed=8969509;
RA   Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA   Miwa Y., Fujita Y.;
RT   "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT   containing the lic and cel loci, and creation of a 177 kb contig
RT   covering the gnt-sacXY region.";
RL   Microbiology 142:3113-3123(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
```

```
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: SOMEHOW INVOLVED IN THE CYTOCHROME D BRANCH OF AEROBIC
CC       RESPIRATION. SEEMS TO BE A COMPONENT OF A TRANSPORT SYSTEM (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MSBA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D83026; BAA11729.1; -.
DR   EMBL; Z99123; CAB15900.1; -.
DR   SubtiList; BG11927; cydC.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   Pfam; PF00005; ABC_tran; 1.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   ProDom; PD000006; ABC_transportr; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport; Transmembrane; Complete proteome.
FT   TRANSMEM     14     34       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM    130    150       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    240    260       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   NP_BIND     360    367       ATP (POTENTIAL).
SQ   SEQUENCE   567 AA;  62806 MW;  74F2500E08C6637D CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 567;
  Best Local Similarity   83.3%;  Pred. No. 27;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 THRIHW 9
           |||:||
Db     519 THRLHW 524


RESULT 9
OBP_HSV11
ID   OBP_HSV11      STANDARD;      PRT;   851 AA.
AC   P10193;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Origin of replication binding protein.
GN   UL9.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88091053; PubMed=2826807;
RA   McGeoch D.J., Dalrymple M.A., Dolan A., McNab D., Perry L.J.,
RA   Taylor P., Challberg M.D.;
RT   "Structures of herpes simplex virus type 1 genes required for
RT   replication of virus DNA.";
RL   J. Virol. 62:444-453(1988).
CC   -!- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC       OF REPLICATION (ORI).
CC   -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC       EHV-1 53, AND VZV 51.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10879; BAA01655.1; -.
DR   EMBL; X14112; CAA32345.1; -.
DR   EMBL; M19120; AAA45822.1; -.
DR   PIR; B29890; WMBEU9.
DR   PIR; I28133; I28133.
DR   TRANSFAC; T00957; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR003450; Herpes_ori_bp.
DR   Pfam; PF02399; Herpes_ori_bp; 1.
DR   SMART; SM00487; DEXDc; 1.
KW   DNA replication; DNA-binding; ATP-binding.
FT   NP_BIND      81     88       ATP (POTENTIAL).
SQ   SEQUENCE   851 AA;  94261 MW;  961A133FE7A30CA7 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 851;
  Best Local Similarity   46.2%;  Pred. No. 41;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESAS 13
             :  |:||:||:
Db     648 STMAARLHWDSAA 660


RESULT 10
YA85_MYCTU
ID   YA85_MYCTU     STANDARD;      PRT;   242 AA.
AC   O53433;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein Rv1085c.
GN   RV1085C OR MT1117 OR MTV017.38C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
```

```
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0073 (HLY-III) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021897; CAA17201.1; -.
DR   EMBL; AE006992; AAK45373.1; -.
DR   TIGR; MT1117; -.
DR   TubercuList; Rv1085c; -.
DR   TIGRFAMs; TIGR01065; hlyIII; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     42     62       POTENTIAL.
FT   TRANSMEM     67     87       POTENTIAL.
FT   TRANSMEM    108    128       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    186    206       POTENTIAL.
FT   TRANSMEM    222    242       POTENTIAL.
SQ   SEQUENCE   242 AA;  26034 MW;  BBC1DE12CF8D3500 CRC64;

  Query Match             48.1%;  Score 38;  DB 1;  Length 242;
  Best Local Similarity   55.6%;  Pred. No. 16;
  Matches    5;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 HRIHWESAS 13
           ||::|:||:
Db      89 HRVNWKSAT 97


RESULT 11
GEM2_HUMAN
ID   GEM2_HUMAN     STANDARD;      PRT;   280 AA.
AC   O14893;
DT   15-JUL-1999 (Rel. 38, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Survival of motor neuron protein-interacting protein 1 (SMN-
DE   interacting protein 1) (Component of gems 2) (Gemin2).
GN   SIP1 OR GEMIN2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast cancer;
RX   MEDLINE=97462902; PubMed=9323129;
RA   Liu Q., Fischer U., Wang F., Dreyfuss G.;
RT   "The spinal muscular atrophy disease gene product, SMN, and its
RT   associated protein SIP1 are in a complex with spliceosomal snRNP
RT   proteins.";
RL   Cell 90:1013-1021(1997).
CC   -!- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICEOSOMAL
CC       SNRNP ASSEMBLY IN THE CYTOPLASM AND IS REQUIRED FOR PRE-MRNA
CC       SPLICING IN THE NUCLEUS.
CC   -!- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR
CC       NEURON PROTEIN (SMN), GEMIN3 AND GEMIN4.
CC   -!- SUBCELLULAR LOCATION: LOCALIZED IN SUBNUCLEAR STRUCTURES NEXT TO
CC       COILED BODIES, CALLED GEMS, WHICH ARE HIGHLY ENRICHED IN
CC       SPLICEOSOMAL SNRNPS. ALSO FOUND IN THE CYTOPLASM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF027150; AAB82297.1; -.
DR   Genew; HGNC:10884; SIP1.
DR   MIM; 602595; -.
KW   mRNA processing; Spliceosome; Nuclear protein.
FT   DOMAIN      101    106       POLY-GLN.
SQ   SEQUENCE   280 AA;  31585 MW;  3232F410EA98EB81 CRC64;

  Query Match             48.1%;  Score 38;  DB 1;  Length 280;
  Best Local Similarity   60.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 HRIHWESASL 14
           || ||:|  |
Db     120 HRSHWKSQQL 129


RESULT 12
YHM7_YEAST
ID   YHM7_YEAST     STANDARD;      PRT;   280 AA.
AC   P38790;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 33.1 kDa protein in SSF1-DYS1 intergenic region.
GN   YHR067W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00061; AAB68378.1; -.
DR   PIR; S46699; S46699.
DR   SGD; S0001109; YHR067W.
KW   Hypothetical protein.
SQ   SEQUENCE   280 AA;  33055 MW;  9FE7B9A602B7A083 CRC64;
```

```
  Query Match             48.1%;  Score 38;  DB 1;  Length 280;
  Best Local Similarity   83.3%;  Pred. No. 19;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 HRIHWE 10
          |||||:
Db    188 HRIHWD 193


RESULT 13
DSCA_HUMAN
ID   DSCA_HUMAN     STANDARD;      PRT;  2012 AA.
AC   O60469; O60468;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Down syndrome cell adhesion molecule precursor (CHD2).
GN   DSCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Brain;
RX   MEDLINE=98087574; PubMed=9426258;
RA   Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA   Lyons G.E., Korenberg J.R.;
RT   "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT   Down syndrome region and is involved in the development of the
RT   nervous system.";
RL   Hum. Mol. Genet. 7:227-237(1998).
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=20384934; PubMed=10925149;
RA   Agarwala K.L., Nakamura S., Tsutsumi Y., Yamakawa K.;
RT   "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT   intercellular adhesion.";
RL   Brain Res. Mol. Brain Res. 79:118-126(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC       INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC       NERVOUS SYSTEM DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC       SHORT ISOFORM MAY BE SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/CHD2-52 (SHOWN HERE)
CC       AND A SHORT FORM/CHD2-42; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023450; AAC17967.1; -.
DR   EMBL; AF023449; AAC17966.1; -.
DR   EMBL; AF217525; AAF27525.1; -.
DR   EMBL; AL163283; CAB90464.1; -.
DR   EMBL; AL163282; CAB90436.1; -.
DR   EMBL; AL163281; CAB90444.1; -.
DR   Genew; HGNC:3039; DSCAM.
DR   MIM; 602523; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 7.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   2012       DOWN SYNDROME CELL ADHESION MOLECULE.
FT   DOMAIN       18   1595       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1596   1616       POTENTIAL.
FT   DOMAIN     1617   2012       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       39    109       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      138    204       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      328    392       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      421    491       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      518    582       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      610    676       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      704    773       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN      802    872       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN      885    972       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      984   1076       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     1088   1177       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     1189   1273       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1300   1366       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN     1380   1463       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1477   1562       FIBRONECTIN TYPE-III 6.
FT   DISULFID     46    102       BY SIMILARITY.
FT   DISULFID    145    197       BY SIMILARITY.
FT   DISULFID    246    293       BY SIMILARITY.
FT   DISULFID    335    385       BY SIMILARITY.
FT   DISULFID    428    484       BY SIMILARITY.
FT   DISULFID    525    575       BY SIMILARITY.
FT   DISULFID    617    669       BY SIMILARITY.
FT   DISULFID    711    766       BY SIMILARITY.
FT   DISULFID    809    865       BY SIMILARITY.
FT   DISULFID   1307   1359       BY SIMILARITY.
FT   CARBOHYD     28     28       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    512    512       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    748    748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    795    795       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    924    924       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1142   1142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1160   1160       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1250   1250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1271   1271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1341   1341       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1488   1488       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1562   1571       NFATLNYDGS -> KEAARCKEFS (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1572   2012       MISSING (IN SHORT ISOFORM).
FT   CONFLICT   1893   2012       HRPGDLIHLPPYLRMDFLLNRGGPGTSRDLSLGQACLEPQK
FT                                SRTLKRPTVLEPIPMEAASSASSTREGQSWQPGAVATLPQR
FT                                EGAELGQAAKMSSSQESLLDSRGHLKGNNPYAKSYTLV ->
FT                                IGQVTSYICLHTLEWTFC (IN REF. 1).
SQ   SEQUENCE   2012 AA;  222259 MW;  0E33CFB781A08334 CRC64;

  Query Match             48.1%;  Score 38;  DB 1;  Length 2012;
  Best Local Similarity   45.5%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 ITHRIHWESAS 13
           :|| :|::| |
Db    1702 VTHTVHYQSVS 1712


RESULT 14
GLMS_YERPE
ID   GLMS_YERPE     STANDARD;      PRT;   608 AA.
AC   Q8Z9S8;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
DE   (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE   phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate
DE   amidotransferase) (Glucosamine-6-phosphate synthase).
GN   GLMS OR YPO4118.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
CC   -!- FUNCTION: Catalyzes the first step in hexosamine metabolism,
CC       converting fructose-6P into glucosamine-6P using glutamine as a
CC       nitrogen source (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-
CC       glutamate + D-glucosamine 6-phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC       GFAT SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ414160; CAC93567.1; -.
DR   InterPro; IPR000583; GATase_2.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF00310; GATase_2; 1.
DR   Pfam; PF01380; SIS; 2.
DR   TIGRFAMs; TIGR01135; glmS; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Transferase; Aminotransferase; Glutamine amidotransferase;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    240       GLUTAMINE AMIDOTRANSFERASE.
FT   ACT_SITE      1      1       GATASE (BY SIMILARITY).
FT   ACT_SITE    603    603       ISOMERIZATION FRU-6P (BY SIMILARITY).
SQ   SEQUENCE   608 AA;  66394 MW;  16E5FD0ADB16CCD6 CRC64;

  Query Match             47.5%;  Score 37.5;  DB 1;  Length 608;
  Best Local Similarity   50.0%;  Pred. No. 53;
  Matches    8;  Conservative    1;  Mismatches    4;  Indels    3;  Gaps    1;

Qy       3 ITHRIHWE---SASLL 15
           | | :|||     |||
Db     127 IAHLVHWEQQQGGSLL 142


RESULT 15
PRT1_PICAN
ID   PRT1_PICAN     STANDARD;      PRT;   220 AA.
AC   P12806;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative PRT1 protein.
GN   PRT1.
OS   Pichia angusta (Yeast) (Hansenula polymorpha).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Pichia.
OX   NCBI_TaxID=4905;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CBS 4732;
RX   MEDLINE=89287321; PubMed=2500147;
RA   Bruinenberg P.G., Evers M., Waterham H.R., Kuipers J., Arnberg A.C.,
RA   Ab G.;
RT   "Cloning and sequencing of the peroxisomal amine oxidase gene from
RT   Hansenula polymorpha.";
RL   Biochim. Biophys. Acta 1008:157-167(1989).
CC   -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15111; CAA33208.1; -.
DR   PIR; S16511; S16511.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   RNA-binding.
FT   DOMAIN       37    120       RNA-BINDING (RRM).
SQ   SEQUENCE   220 AA;  24961 MW;  D317E7EFF49834B5 CRC64;

  Query Match             46.8%;  Score 37;  DB 1;  Length 220;
  Best Local Similarity   58.3%;  Pred. No. 22;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 ITHRIHWESASL 14
           |  | || |||:
Db     184 IEPRAHWTSASM 195


Search completed: February 21, 2003, 14:14:33
Job time : 14 secs
```

OM protein - protein search, using sw model

Run on: February 21, 2003, 14:14:11 ; Search time 29 Seconds
(without alignments)
106.576 Million cell updates/sec

Title: US-09-845-739-1
Perfect score: 79
Sequence: 1 SKITHRIHWESASLL 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	54	68.4	167	6	Q9N0M4	Q9n0m4 cervus nipp
2	54	68.4	349	6	O46544	O46544 ovis aries
3	52	65.8	154	6	Q29289	Q29289 sus scrofa
4	52	65.8	267	16	Q9HTZ5	Q9htz5 pseudomonas
5	52	65.8	1661	6	Q9GKP1	Q9gkp1 sus scrofa
6	46	58.2	441	5	Q8T3J9	Q8t3j9 drosophila
7	45	57.0	211	16	Q9HYZ4	Q9hyz4 pseudomonas
8	42	53.2	401	16	Q9P9Y5	Q9p9y5 xylella fas
9	42	53.2	407	5	Q8SYY7	Q8syy7 drosophila
10	42	53.2	411	5	Q9V4I4	Q9v4i4 drosophila
11	41	51.9	197	17	Q9HK18	Q9hk18 thermoplasm
12	41	51.9	219	13	Q90YC5	Q90yc5 brachydanio
13	41	51.9	229	16	Q8YIY6	Q8yiy6 brucella me
14	41	51.9	318	2	Q9X5J4	Q9x5j4 mycobacteri
15	41	51.9	336	16	Q9RX07	Q9rx07 deinococcus
16	41	51.9	386	2	Q9AEX8	Q9aex8 treponema h
17	41	51.9	541	16	Q9A017	Q9a017 streptococc
18	41	51.9	615	16	Q9CHM3	Q9chm3 lactococcus
19	40.5	51.3	1417	16	Q8X6G3	Q8x6g3 escherichia
20	40	50.6	205	5	Q9NDY6	Q9ndy6 leishmania
21	40	50.6	232	16	Q92KX1	Q92kx1 rhizobium m
22	40	50.6	272	16	Q984A5	Q984a5 rhizobium l
23	40	50.6	274	11	Q9D912	Q9d912 mus musculu
24	40	50.6	285	5	Q18611	Q18611 caenorhabdi
25	40	50.6	332	2	O82937	O82937 escherichia
26	40	50.6	343	2	Q9ZGU3	Q9zgu3 escherichia
27	40	50.6	420	5	Q9VRZ4	Q9vrz4 drosophila
28	40	50.6	1145	11	Q9DBV3	Q9dbv3 mus musculu
29	39.5	50.0	102	11	Q8R352	Q8r352 mus musculu
30	39	49.4	75	6	Q9GMH7	Q9gmh7 macaca fasc
31	39	49.4	191	12	Q9E348	Q9e348 maize necro
32	39	49.4	266	16	Q8UBV4	Q8ubv4 agrobacteri
33	39	49.4	321	11	Q91ZC0	Q91zc0 mus musculu
34	39	49.4	322	11	Q91ZB9	Q91zb9 mus musculu
35	39	49.4	381	2	Q8RTQ7	Q8rtq7 thermodesul
36	39	49.4	382	2	Q93EV7	Q93ev7 thermodesul
37	39	49.4	406	2	Q9JPB4	Q9jpb4 rhodocyclus
38	39	49.4	406	2	P95619	P95619 rhodocyclus
39	39	49.4	454	16	Q987Y3	Q987y3 rhizobium l
40	39	49.4	574	16	Q927C5	Q927c5 listeria in
41	39	49.4	574	16	Q8Y3W3	Q8y3w3 listeria mo
42	38	48.1	193	4	Q96DJ2	Q96dj2 homo sapien
43	38	48.1	215	17	Q97AJ2	Q97aj2 thermoplasm
44	38	48.1	224	16	Q98FI2	Q98fi2 rhizobium l
45	38	48.1	248	16	Q8ZJK6	Q8zjk6 yersinia pe

ALIGNMENTS

RESULT 1
Q9N0M4
ID Q9N0M4 PRELIMINARY; PRT; 167 AA.
AC Q9N0M4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Complement C3 alpha chain (Fragment).
OS Cervus nippon (Sika deer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidea;
OC Cervidae; Cervinae; Cervus.
OX NCBI_TaxID=9863;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jiang Y., Sun L.G., Yu Y.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF264631; AAF73464.1; -.
DR HSSP; P01024; 1C3D.
DR InterPro; IPR001599; MacroglobInA2.
DR Pfam; PF00207; A2M; 1.
FT NON_TER 1 1
SQ SEQUENCE 167 AA; 18671 MW; 12BFE0798290DFA7 CRC64;

Query Match 68.4%; Score 54; DB 6; Length 167;
Best Local Similarity 73.3%; Pred. No. 0.099;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 SKITHRIHWESASLL 15
           | : ||| |||||||
Db      48 SLVKHRILWESASLL 62
```

RESULT 2
O46544
ID O46544 PRELIMINARY; PRT; 349 AA.
AC O46544;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)

```
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Complement component C3 (Fragment).
GN   C3.
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE ALPINE; TISSUE=LIVER;
RX   MEDLINE=98309471; PubMed=9647256;
RA   Hein W.R., Dudler L., Marston W.L., Landsverk T., Young A.J.,
RA   Avila D.;
RT   "Ubiquitination and dimerization of complement receptor type 2 on
RT   sheep B cells.";
RL   J. Immunol. 161:458-466(1998).
DR   EMBL; AF038130; AAB92374.2; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   Pfam; PF00207; A2M; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
FT   NON_TER       1      1
FT   NON_TER     349    349
SQ   SEQUENCE   349 AA;  39679 MW;  70C2023E42ED5EE3 CRC64;

  Query Match             68.4%;  Score 54;  DB 6;  Length 349;
  Best Local Similarity   73.3%;  Pred. No. 0.22;
  Matches   11;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          | : ||| |||||||
Db    329 SLVKHRILWESASLL 343


RESULT 3
Q29289
ID   Q29289      PRELIMINARY;      PRT;   154 AA.
AC   Q29289;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Complement C3 (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SMALL INTESTINE;
RX   MEDLINE=96327607; PubMed=8672129;
RA   Winteroe A.K., Fredholm M., Davies W.;
RT   "Evaluation and characterization of a porcine small intestine cDNA
RT   library.";
RL   Mamm. Genome 7:509-517(1996).
DR   EMBL; F14640; CAA23173.1; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   Pfam; PF00207; A2M; 1.
FT   NON_TER       1      1
FT   NON_TER     154    154
SQ   SEQUENCE   154 AA;  17440 MW;  6DC7661C1253ED45 CRC64;

  Query Match             65.8%;  Score 52;  DB 6;  Length 154;
  Best Local Similarity   76.9%;  Pred. No. 0.2;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 ITHRIHWESASLL 15
          : ||| |||||||
Db    100 VRHRILWESASLL 112


RESULT 4
Q9HTZ5
ID   Q9HTZ5      PRELIMINARY;      PRT;   267 AA.
AC   Q9HTZ5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA5194.
GN   PA5194.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004932; AAG08579.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   267 AA;  30527 MW;  57CD9D2319B6AD7D CRC64;

  Query Match             65.8%;  Score 52;  DB 16;  Length 267;
  Best Local Similarity   57.1%;  Pred. No. 0.37;
  Matches    8;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASL 14
          :|| | :||: |||
Db    118 AKIAHHLHWQHASL 131


RESULT 5
Q9GKP1
ID   Q9GKP1      PRELIMINARY;      PRT;  1661 AA.
AC   Q9GKP1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Complement component C3.
GN   C3.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE=21313047; PubMed=11419349;
RA   Wimmers K., Mekchay S., Ponsuksili S., Hardge T., Yerle M.,
RA   Schellander K.;
RT   "Polymorphic sites in exon 15 and 30 of the porcine C3 gene.";
RL   Anim. Genet. 32:46-47(2001).
DR   EMBL; AF154933; AAG40565.1; -.
DR   HSSP; P01024; 1C3D.
DR   InterPro; IPR002890; A2M_N.
DR   InterPro; IPR000020; Anaphylatoxin.
DR   InterPro; IPR001840; Anaphylatoxn.
DR   InterPro; IPR001599; MacrogloblnA2.
DR   InterPro; IPR001134; Netrin_C.
DR   Pfam; PF00207; A2M; 1.
DR   Pfam; PF01835; A2M_N; 1.
DR   Pfam; PF01821; ANATO; 1.
DR   Pfam; PF01759; NTR; 1.
```

```
DR   PRINTS; PR00004; ANAPHYLATOXN.
DR   ProDom; PD003264; Anaphylatoxin; 1.
DR   SMART; SM00104; ANATO; 1.
DR   PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.
DR   PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR   PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
SQ   SEQUENCE   1661 AA;  186806 MW;  4899D0914BE3310C CRC64;

  Query Match             65.8%;  Score 52;  DB 6;  Length 1661;
  Best Local Similarity   76.9%;  Pred. No. 2.6;
  Matches   10;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 ITHRIHWESASLL 15
          : ||| |||||||
Db   1305 VRHRILWESASLL 1317


RESULT 6
Q8T3J9
ID   Q8T3J9      PRELIMINARY;      PRT;   441 AA.
AC   Q8T3J9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   AT11889p.
GN   CG7196.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY094997; AAM11325.1; -.
SQ   SEQUENCE   441 AA;  52125 MW;  847067D8FA3A3A16 CRC64;

  Query Match             58.2%;  Score 46;  DB 5;  Length 441;
  Best Local Similarity   50.0%;  Pred. No. 7;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 KITHRIHWESASLL 15
          |: |: ||   |||
Db     20 KVVHKNHWRQVSLL 33


RESULT 7
Q9HYZ4
ID   Q9HYZ4      PRELIMINARY;      PRT;   211 AA.
AC   Q9HYZ4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Pseudouridine synthase RluA.
GN   RLUA OR PA3246.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004747; AAG06634.1; -.
DR   InterPro; IPR000613; PseudoU_synth.
DR   InterPro; IPR002990; PSI_RLU.
DR   Pfam; PF00849; PseudoU_synth_2; 1.
DR   ProDom; PD001819; PseudoU_synth; 1.
DR   PROSITE; PS01129; PSI_RLU; 1.
KW   Complete proteome.
SQ   SEQUENCE   211 AA;  24338 MW;  D333B20FCEA55A94 CRC64;

  Query Match             57.0%;  Score 45;  DB 16;  Length 211;
  Best Local Similarity   40.0%;  Pred. No. 4.8;
  Matches    6;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWESASLL 15
          ::| ||: ||:: |:
Db     50 ARIVHRLDWETSGLM 64


RESULT 8
Q9P9Y5
ID   Q9P9Y5      PRELIMINARY;      PRT;   401 AA.
AC   Q9P9Y5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Xf2735.
GN   XF2735.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004080; AAF85520.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   401 AA;  45544 MW;  050ADA91253A6398 CRC64;

  Query Match             53.2%;  Score 42;  DB 16;  Length 401;
```

```
  Best Local Similarity   45.5%;  Pred. No. 32;
  Matches    5;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 ITHRIHWESAS 13
           : ||:||:  |
Db     334 LAHRVHWDEES 344


RESULT 9
Q8SYY7
ID   Q8SYY7      PRELIMINARY;      PRT;   407 AA.
AC   Q8SYY7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   RE27547p.
GN   CG1859.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY071238; AAL48860.1; -.
SQ   SEQUENCE   407 AA;  44863 MW;  5D2A46A75CB6DD78 CRC64;

  Query Match             53.2%;  Score 42;  DB 5;  Length 407;
  Best Local Similarity   77.8%;  Pred. No. 32;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 HRIHWESAS 13
           ||: |||||
Db     150 HRLSWESAS 158


RESULT 10
Q9V4I4
ID   Q9V4I4      PRELIMINARY;      PRT;   411 AA.
AC   Q9V4I4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   CG1859 protein.
GN   CG1859.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR   EMBL; AE003842; AAF59286.1; -.
DR   FlyBase; FBgn0033147; CG1859.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00079; serpin; 1.
DR   SMART; SM00093; SERPIN; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin.
SQ   SEQUENCE   411 AA;  45275 MW;  8B5DEB6340C468A3 CRC64;

  Query Match             53.2%;  Score 42;  DB 5;  Length 411;
  Best Local Similarity   77.8%;  Pred. No. 33;
  Matches    7;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 HRIHWESAS 13
           ||: |||||
Db     150 HRLSWESAS 158


RESULT 11
Q9HK18
ID   Q9HK18      PRELIMINARY;      PRT;   197 AA.
AC   Q9HK18;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Endonuclease III related protein.
GN   TA0790.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC   Thermoplasmataceae; Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445065; CAC11921.1; -.
DR   HSSP; P20625; 2ABK.
DR   InterPro; IPR004035; EndoIII_FCL.
DR   InterPro; IPR003265; Endo_3c.
DR   InterPro; IPR003651; FeS_bind.
```

```
DR   InterPro; IPR003583; HHH_1.
DR   Pfam; PF00730; HhH-GPD; 1.
DR   SMART; SM00478; ENDO3c; 1.
DR   SMART; SM00525; FES; 1.
DR   SMART; SM00278; HhH1; 1.
DR   PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   197 AA;  22753 MW;  E00E2162AD95A856 CRC64;

  Query Match             51.9%;  Score 41;  DB 17;  Length 197;
  Best Local Similarity   63.6%;  Pred. No. 22;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 KITHRIHWESA 12
          :|:||| | ||
Db    125 RISHRIGWSSA 135


RESULT 12
Q90YC5
ID   Q90YC5      PRELIMINARY;      PRT;   219 AA.
AC   Q90YC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Ephrin-A3.
GN   EPHRIN-A3.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21412237; PubMed=11520665;
RA   Hirate Y., Mieda M., Harada T., Yamasu K., Okamoto H.;
RT   "Identification of ephrin-A3 and novel genes specific to the midbrain-
RT   MHB in embryonic zebrafish by ordered differential display.";
RL   Mech. Dev. 107:83-96(2001).
DR   EMBL; AB051678; BAB55891.1; -.
DR   InterPro; IPR001799; Ephrin.
DR   Pfam; PF00812; Ephrin; 1.
DR   ProDom; PD002533; Ephrin; 1.
DR   PROSITE; PS01299; EPHRIN; UNKNOWN_1.
SQ   SEQUENCE   219 AA;  25146 MW;  7191927E03F8EA01 CRC64;

  Query Match             51.9%;  Score 41;  DB 13;  Length 219;
  Best Local Similarity   45.5%;  Pred. No. 25;
  Matches    5;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      5 HRIHWESASLL 15
          | :|| |:::|
Db     26 HAVHWNSSNIL 36


RESULT 13
Q8YIY6
ID   Q8YIY6      PRELIMINARY;      PRT;   229 AA.
AC   Q8YIY6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical cytosolic protein BMEI0303.
GN   BMEI0303.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR   EMBL; AE009472; AAL51484.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   229 AA;  25507 MW;  98A1769A370F52CA CRC64;

  Query Match             51.9%;  Score 41;  DB 16;  Length 229;
  Best Local Similarity   53.8%;  Pred. No. 26;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 KITHRIHWESASL 14
          :| :| || ||:|
Db    136 QIRNRTHWNSANL 148


RESULT 14
Q9X5J4
ID   Q9X5J4      PRELIMINARY;      PRT;   318 AA.
AC   Q9X5J4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   Hemolytic protein hlpA.
GN   HLPA.
OS   Mycobacterium avium.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1764;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TMC 724;
RA   Eckstein T.M., Brennan P.J., Inamine J.M., Belisle J.T.;
RT   "Identification of gene cluster involved in glycopeptidolipid
RT   biosynthesis and of a gene cluster encoding daunorubicin resistance in
RT   two strains of Mycobacterium avium serovar 2.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125999; AAD20370.1; -.
SQ   SEQUENCE   318 AA;  37178 MW;  C153D903897BEF35 CRC64;

  Query Match             51.9%;  Score 41;  DB 2;  Length 318;
  Best Local Similarity   70.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 SKITHRIHWE 10
          || | | |||
Db    187 SKTTERYHWE 196


RESULT 15
Q9RX07
ID   Q9RX07      PRELIMINARY;      PRT;   336 AA.
AC   Q9RX07;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MRR restriction system protein.
GN   DR0508.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
```

```
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001910; AAF10088.1; -.
DR   TIGR; DR0508; -.
KW   Complete proteome.
SQ   SEQUENCE   336 AA;  37335 MW;  E978C50EC4BBC17B CRC64;

  Query Match             51.9%;  Score 41;  DB 16;  Length 336;
  Best Local Similarity   50.0%;  Pred. No. 39;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 SKITHRIHWESASL 14
           ||: ||| |  ::|
Db      72 SKVRHRIAWACSNL 85
```

Search completed: February 21, 2003, 14:16:41
Job time : 35 secs